

Ergebnisse der Methodenstudie zur internationalen Vergleichbarkeit von Einstellungsskalen in der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1982

Faulbaum, Frank

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Faulbaum, F. (1984). *Ergebnisse der Methodenstudie zur internationalen Vergleichbarkeit von Einstellungsskalen in der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1982*. (ZUMA-Arbeitsbericht, 1984/04). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-66306>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Frank Faulbaum

Ergebnisse der Methodenstudie zur internationalen
Vergleichbarkeit von Einstellungsskalen in der
Allgemeinen Bevölkerungsumfrage der Sozialwissen-
schaften (ALLBUS) 1982

Mannheim, Juli 1984

ZUMA-Arbeitsbericht Nr. 04/84

Ab Juli 1983 sind die bisherigen ZUMA-Arbeitsberichte in zwei Reihen aufgeteilt:

Die ZUMA-Arbeitsberichte (neue Folge) haben eine hausinterne Begutachtung durchlaufen und werden vom Geschäftsführenden Direktor zusammen mit den übrigen wissenschaftlichen Leitern herausgegeben. Die Berichte dieser Reihe sind zur allgemeinen Weitergabe nach außen bestimmt.

Die ZUMA-Technischen Berichte dienen dem Zweck der hausinternen Kommunikation bzw. der Unterrichtung externer Kooperationspartner. Sie sind nicht zur allgemeinen Weitergabe bestimmt.

•

INHALT

	<u>Seite</u>
VORWORT	
1. PROBLEMSTELLUNG UND FORSCHUNGSPROGRAMM	1
1.1 Das Problem der Vergleichbarkeit	1
1.2 Das ursprüngliche Forschungsprogramm	4
1.3 Das realisierte Forschungsprogramm	8
2. PRAKTISCHE DURCHFÜHRUNG	10
2.1 Das Meßinstrument	10
2.2 Stichproben	14
2.2.1 ZUMA-Studie	14
2.2.2 NORC-Studie	16
2.2.3 Vergleichende Betrachtung einiger zentraler demographischer Variablen	17
3. ERGEBNISSE DER SEMANTISCHEN UNTERSUCHUNGEN	20
3.1 Einleitende Bemerkungen	20
3.2 Modelle zur Vorhersage und Erklärung der Wirkungsstärken adverbialer Modifikatoren	22
3.3 Zusammenfassende Darstellung der Befunde	24
4. ERGEBNISSE DES VERGLEICHS ZWISCHEN NUMERISCHEN VERBALEN KATEGORIALSKALEN	27
4.1 Einleitende Bemerkungen	27
4.2 Betrachtung der Einzelitems	28
4.3 Identifikation zugrundeliegender Konstrukte	31
4.3.1 Ergebnisse der exploratorischen Faktorenanalysen	31
4.3.2 Identifikation von Indikatoren	34
4.4 Konstruktinvarianz numerischer und verbaler Skalen	41
5. DIE INTERNATIONALE VERGLEICHBARKEIT DER BERUFSWERTE-SKALA	48
5.1 Vorbemerkungen	48
5.2 Die vergleichende Beurteilung der meßtheoretischen Qualitäten	48
5.2.1 Die Magnitude-Messungen	48
5.2.2 Die kategorialen Messungen	57

INHALT (FORTSETZUNG)

	<u>Seite</u>
5.3 Konstruktidentifikation und Konstruktvergleiche	58
5.3.1 Faktoren und Faktorenstrukturen im Vergleich	58
5.3.2 Ergebnisse der PINDIS-Analysen	65
5.4 Die Identifikation von Indikatoren	75
5.5 Zur Vorhersage von Gruppenunterschieden	80
6. ZUSAMMENFASSUNG UND SCHLUSSFOLGERUNGEN	83
LITERATUR	

VORWORT

Der vorliegende Forschungsbericht enthält eine Darstellung der Ergebnisse der methodischen Zusatzstudie zum ALLBUS 1982, die sich mit der internationalen Vergleichbarkeit von Einstellungsskalen beschäftigte. Verglichen wurden eine Teilstichprobe des vom National Opinion Research Center (NORC) der Universität Chicago betreuten General Social Survey (GSS) 1982 und eine Zusatzstichprobe zur von ZUMA betreuten Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1982 hinsichtlich der Wichtigkeitsskalierungen von Berufswerten. Das Forschungsvorhaben wurde von der Deutschen Forschungsgemeinschaft finanziell unterstützt (Aktenzeichen LE 298/4-5). Antragsteller war Professor M. Rainer Lepsius, Universität Heidelberg, gemeinsam mit Professor Erwin K. Scheuch, Universität Köln, und Professor Rolf Ziegler, Universität München. Die Untersuchungsanlage wurde von Bernd Wegener und Frank Faulbaum (beide ZUMA) konzipiert.

Verschiedene Aspekte der Studie wurden bereits in mehreren Aufsätzen und Berichten dargestellt (vgl. WEGENER 1982a, 1983a; FAULBAUM 1983, 1984; WEGENER, FAULBAUM & MAAG 1982a, 1982b). Der hier vorgelegte Ergebnisbericht stellt einen Kompromiß dar zwischen dem Bedürfnis nach einer möglichst vollständigen Ergebnispräsentation einerseits und dem Bedürfnis nach möglichst geringer Redundanz andererseits. So werden die bereits an anderer Stelle publizierten Ergebnisse zwar dargestellt; dies geschieht aber in einer etwas geraffteren Form. In ähnlicher Weise wird auch mit der Darstellung der angewandten Methoden und Analysetechniken verfahren. So ist z.B. über Theorie und Technik der Magnitudeskalierung bereits an anderer Stelle mehrfach berichtet worden (vgl. z.B. WEGENER 1978, 1980), sodaß keine zwingende Notwendigkeit bestand, sie noch einmal in extenso vorzustellen. Das gleiche gilt für einige Auswertungstechniken wie z.B. die Anwendung von Strukturgleichungsmodellen. Auch hier wird teilweise einfach auf die existierende Literatur verwiesen.

1. PROBLEMSTELLUNG UND FORSCHUNGSPROGRAMM

1.1 Das Problem der Vergleichbarkeit

Neben organisatorischen Problemen sind es vor allem methodologische Probleme, die in Projekten international vergleichender Umfrageforschung bewältigt werden müssen (vgl. ALLERBECK 1977; BRISLIN 1983; BRISLIN et al. 1973; ELDER 1976; NAROLL 1968; NOWAK 1977; PRZEWORSKI & TEUNE 1966, 1970; SCHEUCH 1967, 1968; TEUNE 1968, 1977; WHITING 1968). Einige dieser Probleme betreffen in besonderer Weise die vergleichende Einstellungsmessung, in deren Bereich die vorliegende Studie anzusiedeln ist. Vergleichende Einstellungsmessungen gehören zum Standardrepertoire großer Vergleichsstudien wie der Partizipations-Studie von VERBA et al. (vgl. z.B. VERBA 1977; VERBA et al. 1978), der Werte-Studie von JACOB et al. (vgl. z.B. JACOB & JACOB 1977) oder der 'Political Action'-Studie von BARNES, KAASE et al. (vgl. BARNES, KAASE et al. 1979). Auch im Fall des ALLBUS 1982 und des GSS 1982 wurde der Versuch unternommen, durch eine teilweise Angleichung des Fragenprogramms die Voraussetzungen für vergleichende Einstellungsmessungen zu schaffen.

Die zunächst durch bloße Sprachübersetzung erzielte Obereinstimmung im Fragenprogramm ist allerdings erst eine der Voraussetzungen für die Durchführung vergleichender Einstellungsmessungen. Eine weitere, ebenso wichtige Voraussetzung ist die Vergleichbarkeit der verwendeten, in der Regel aus mehreren Items zusammengesetzten Einstellungsskalen. Vergleichbarkeit bedeutet, daß die Skalen in den jeweils betrachteten Nationen identische Konstrukte messen (Validitätsproblem) und daß diese Messungen vergleichbar reliabel sind (Reliabilitätsproblem).

Der Nachweis der Vergleichbarkeit ist keineswegs trivial und bedarf eines besonderen methodischen Vorgehens. Ziel der methodischen Zusatzstudie zum ALLBUS 1982 war es, paradigmatisch zu demonstrieren, wie eine Strategie zur Überprüfung der Vergleichbarkeit aussehen könnte.

Um den finanziellen und personellen Aufwand der Studie in praktikablem Rahmen zu halten, wurde beschlossen, diese Problemstellung am Beispiel

einer einzelnen, dem Fragenprogramm des ALLBUS 1982 und des GSS 1982 gemeinsamen Itembatterie zu untersuchen. Diese Itembatterie war die zur subjektiven Wichtigkeit von Merkmalen beruflicher Arbeit (Arbeitsorientierungen; ALLBUS 1982: Frage 6A; GSS 1982: Frage 84Y). Sie enthielt die in Tabelle 1 dargestellten 13 deutschen bzw. amerikanischen Items. Der genaue Fragetext lautete:

- Für wie wichtig halten Sie persönlich diese Merkmale für die berufliche Arbeit und den Beruf?
(How important do you personally consider the following job characteristics?)

Tab. 1: Itembatterie der Berufswerte +

A. Sichere Berufsstellung (SICHER)	A. Job security (KEEP)
B. Hohes Einkommen (EINKOM)	B. High income (PAY)
C. Gute Aufstiegsmöglichkeiten (AUFSTG)	C. Good opportunities for advancement (RISE)
D. Ein Beruf, der anerkannt und geachtet wird (GEACHT)	D. An occupation that is recognized and respected (HONR)
E. Ein Beruf, der einem viel Freizeit läßt (FRZEIT)	E. An occupation that leaves one a lot of leisure time (OFF)
F. Interessante Tätigkeit (INTER)	F. An interesting job (INTE)
G. Eine Tätigkeit, bei der man selbständig arbeiten kann (SELBST)	G. A job that allows one to work independently (INDP)
H. Aufgaben, die viel Verantwortungsbewußtsein erfordern (VERANT)	H. Responsible job tasks (RESP)
I. Viel Kontakt zu anderen Menschen (KONTKT)	I. A lot of contact with other people (PEOP)
K. Ein Beruf, bei dem man anderen helfen kann (HELFEN)	K. An occupation in which one can help others (HELP)
L. Ein Beruf, der für die Gesellschaft nützlich ist (NUETZL)	L. A job that is useful to society (SOC)
M. Gibt mir das Gefühl, etwas Sinnvolles zu tun (SINNVL)	M. Gives the feeling of doing something meaningful (ACMP)
N. Sichere und gesunde Arbeitsbedingungen (BEDING)	N. Safe and healthy working conditions (SAFE)

+ Die Angaben in Klammern stellen die im Text verwendeten Abkürzungen dar.

Die Herstellung der Vergleichbarkeit von Einstellungsmessungen ist normalerweise ein Vorgang, der vor dem Einsatz in einer Hauptuntersuchung erfolgen sollte und der vor dem Problem steht, für die ins Auge gefaßten Konstrukte in den verschiedenen Nationen funktional äquivalente Indikatoren zu finden (vgl. hierzu z.B. SCHEUCH 1967:176; ALLERBECK 1977:378; VERBA et al. 1978: 36). Das methodische Vorgehen ist dabei ein mehr deduktionistisches: Man geht von einem vorgegebenen theoretischen Begriff aus, der dann in mehreren Nationen - durch möglicherweise unterschiedliche Indikatoren - operationalisiert wird, wobei die Indikatoren erst gefunden werden müssen.

Im vorliegenden Fall der Methodenstudie war der Nachweis der Vergleichbarkeit nur post hoc und auf eher induktionistischem Wege möglich, wobei mittels formaler statistischer Analysen aus den Befragtenreaktionen in Bezug auf eine vorgegebene Menge von Items auf eine zugrundeliegende Menge von Konstrukten mit ihren Indikatoren geschlossen werden mußte. Es liegt auf der Hand, daß ein solches Vorgehen zunächst nur zur Aufdeckung international ähnlicher Reaktionsmuster führen kann, wobei letztere nur Hinweise auf zugrundeliegende Konstrukte geben können. Die Identifizierung von Konstrukten ist ein Interpretationsprozeß, der den Einsatz zusätzlichen theoretischen Hintergrundwissens und gegebenenfalls zusätzliche empirische Untersuchungen erfordert, bei denen theoretisch abgeleitete Beziehungen zwischen den vermeintlichen theoretischen Konstrukten und weiteren theoretisch relevanten Variablen überprüft werden. Gerade bei der post hoc-Identifikation von Konstrukten wird die Forderung TEUNES (TEUNE 1977:99) relevant, daß man theoretischen Überlegungen mehr Aufmerksamkeit schenken sollte als der 'Mechanik der Datengewinnung'.

Eine stringente Überprüfung der Vergleichbarkeit erfordert streng genommen die folgenden Schritte:

1. Überprüfung der semantischen Invarianz der Itemformulierungen

Da Befragte nicht auf die sprachliche Oberflächenstruktur der Itemformulierungen reagieren, sondern auf die kognitiv-semantische Repräsentation in ihrem kognitiven System, müßte man streng genommen überprüfen, ob eine in der Sprache L_1 erfolgte Formulierung eines Items bei Befragten der Nation 1 die gleiche semantische Repräsentation erzeugt wie die Formu-

lierung des gleichen Items in der Sprache L_2 bei Befragten der Nation 2, oder ob es systematische Unterschiede in den semantischen Repräsentationen gibt. Dieses Problem tritt in abgeschwächter Form auch in Ein-Nationen-Studien auf.

2. Identifikation der zugrundeliegenden Konstrukte und ihrer Indikatoren

Der zweite Schritt besteht in dem Versuch, den Itemmeßwerten zugrundeliegende Konstrukte mit ihren Indikatoren zu identifizieren, d.h. in der Identifikation eventuell zugrundeliegender Meßmodelle. Außerdem umfaßt dieser Schritt auch die Bestimmung der Reliabilitäten, mit denen die gefundenen Indikatoren die ihnen zugeordneten Konstrukte messen.

Werden die Items auf verbal abgestuften Skalen beurteilt, so wird ein weiterer Zwischenschritt zur Überprüfung der semantischen Invarianz der verbalen Skalenbenennungen erforderlich.

1.2 Das ursprüngliche Forschungsprogramm

Was den ersten Schritt des im letzten Abschnitt geschilderten Vorgehens zum Nachweis der Vergleichbarkeit (Überprüfung der semantischen Invarianz der Itemformulierungen) angeht, so fällt eine experimentelle Überprüfung in völlig andere Forschungskontexte, nämlich die Forschungskontexte der Psycho- und Soziolinguistik und der kognitiven Psychologie mit ihren z.Teil in Umfragen nicht ohne weiteres anwendbaren experimentellen Methoden. Um das Problem der semantischen Invarianz zu entschärfen, haben wir stattdessen darauf geachtet, daß die Übersetzung der Items in enger Zusammenarbeit mit dem amerikanischen Forschungspartner erfolgte. Allerdings wurde beschlossen, einen Teilaspekt des Invarianzproblems paradigmatisch an einem praktikablen Beispiel zu untersuchen. Das gewählte Beispiel betraf die subjektiven Intensitätseinstufungen adverbial modifizierter Adjektive, wie sie in Einstellungsskalen häufig innerhalb von Itemformulierungen auftreten. Im vorliegenden Beispiel waren es das Adjektiv 'sicher', das in dem Item 'Sichere Berufsstellung' auftritt, sowie seine Modifikationen durch die Adverbien

'kaum', 'wenig', 'etwas', 'teilweise', 'einigermaßen', 'halbwegs',
'überwiegend', 'ziemlich', 'sehr', 'außerordentlich'.

Was den zweiten Schritt des im letzten Abschnitt genannten Verfahrens zur Vergleichbarkeitsüberprüfung betrifft, so war geplant, drei alternative Skalierungsarten in ihren Auswirkungen auf die Vergleichbarkeit von Einstellungsmessungen zu überprüfen:

1. Die auch im ALLBUS verwendete 7-stufige numerische Kategorienskala zur Wichtigkeitseinstufung der Berufsmerkmale, bei der lediglich die Endpunkte verbal benannt waren (1='unwichtig' bzw. 'not important at all', 0='sehr wichtig' bzw. 'very important').
2. Eine verbale Version der 7-stufigen Kategorienskala, wobei die 7 Abstufungen in folgender Weise verbal benannt wurden:

'nicht wichtig'(1), 'wenig wichtig'(2), 'etwas wichtig'(3),
'einigermaßen wichtig'(4), 'ziemlich wichtig'(5), 'überwiegend wichtig'(6) und 'sehr wichtig'(7).

Die Adverbien 'nicht', 'etwas' etc. hatten sich in einer Untersuchung von ROHRMANN (1978) als relativ gleichabständig herausgestellt.

3. Eine bi-modale Magnitude-Skalierung der Berufsmerkmale (Modalitäten: Linien und Zahlen).

Die Hineinnahme der Magnitude-Skalierung hatte allerdings noch eine wichtigere Funktion: Sie sollte als Kontrollskalierung für die Kategorienskalisierung dienen. In einer großen Anzahl von Untersuchungen im Rahmen der ZUMA-Forschungen zur Methodenentwicklung konnte gezeigt werden, daß zwischen kategorischen Messungen und Magnitude-Messungen ganz bestimmte mathematische Beziehungen (Interskalenrelationen) angenommen werden müssen, deren Anpassungsgüten über die Magnitude-Skalierung eine Abschätzung der Leistungsfähigkeit von kategorischen Messungen erlauben (vgl. z.B. WEGENER & KIRSCHNER 1982; WEGENER 1983b).

Ein weiterer Vorteil der Magnitude-Skalierung besteht darin, daß ihr eine explizite, psychophysisch begründbare Urteilstheorie zugrundeliegt, die bestimmte Annahmen über die Art der individuellen Urteilsfunktionen macht. Unter einer Urteilsfunktion versteht man dabei die funktionale Beziehung zwischen der Intensität der "wahren" Empfindungswerte und den Ausprägungen beobachtbarer Reaktionen, mit denen Befragte diese Empfindungsintensitäten ausdrücken. Verschiedene Autoren (vgl. SARIS et al. 1980; WEGENER 1983b) haben gezeigt, wie man die Parameter dieser Urteilsfunktionen im Rahmen konfirmatorischer Faktorenanalysen bestimmen kann. Dabei werden die Urteilsfunktionen als 'Pfade' von den latenten Konstrukten auf die Indikatoren (Linien, Zahlen) abgebildet. Voraussetzungen für diese Analysen sind einmal eine genügend große Anzahl von Stimuli und das Vorhandensein von mindestens zwei latenten Konstrukten.

In der vorliegenden Untersuchung war geplant, durch die Bestimmung der individuellen Urteilsfunktionen der Befragten beider Nationen zu überprüfen, inwiefern nationale Unterschiede in den Befragtenreaktionen auf unterschiedliche Urteilsfunktionen zurückzuführen sind. Die Stimuli sollten dabei die 13 Berufsmerkmale sein. Als Konstrukte waren vorgesehen die Wichtigkeitseinstufung in bezug auf den Beruf allgemein einerseits und das Ausmaß, in dem die Berufsmerkmale im eigenen Beruf verwirklicht sind, andererseits.

Die Hineinnahme einer verbalen Version der kategorialen Wichtigkeitsskala ließ in Übereinstimmung mit der in Abschnitt 1.1 erläuterten Strategie die Bestimmung der subjektiven Intensitäten der verbalen Abstufungen sinnvoll erscheinen, um gegebenenfalls abschätzen zu können, in wieweit nationale Unterschiede in den Itemmeßwerten auf Unterschiede des semantischen Verständnisses der verbalen Skalenbenennungen zurückzuführen sind. Im Zusammenhang mit der Bestimmung der subjektiven Intensitäten des Adjektivs 'sicher' in dem Item 'Sichere Berufsstellung' hätte sich zugleich die Gelegenheit ergeben festzustellen, inwiefern sich Wechselwirkungen zwischen dem len Verständnis von 'sicher' und dem individuellen Verständnis der verbalen Abstufungen des Adjektivs 'wichtig' auf die Itemmeßwerte auswirken.

Die vollständige Realisierung des bisher beschriebenen Forschungsprogramms hätte die Aufnahme der folgenden Aufgaben in ein geeignetes Meßinstrument erforderlich gemacht:

1. Kategoriale Einstufungen der 13 Berufswerte nach ihrer subjektiven Wichtigkeit auf einer 7-stufigen numerischen Kategorialskala.
2. Kategoriale Einstufungen der Berufswerte auf einer 7-stufigen verbalen Version der Wichtigkeitsskala, bei der die kategorialen Abstufungen wie auf S. 5 angegeben benannt sind.
3. Bimodale Magnitude-Einschätzungen der Berufswerte hinsichtlich ihrer subjektiven Wichtigkeit mit Linien und Zahlen.
4. Bi-modale Magnitude-Einschätzungen (Linien, Zahlen) der unmodifizierten und modifizierten Adjektive 'sicher' und 'wichtig' hinsichtlich des 'Ausmaßes an Sicherheit' bzw. 'Ausmaßes an Wichtigkeit', das durch sie für die Befragten ausgedrückt wird, um so die Unterschiede im Verständnis des Stimulus 'Sichere Berufsstellung' als auch Unterschiede im Verständnis der Antwortkategorien der verbalen Kategorialskala zu erfassen. Um einen Überblick darüber zu erhalten, inwieweit die Wirkung adverbialer Modifikatoren vom gewählten Adjektiv abhängt, wurde das Adjektiv 'wichtig' auch in Zusammenhang mit jenen Modifikatoren eingestuft, die nicht zur Menge der verbalen Abstufungen der verbalen Version der Wichtigkeitsskala, wohl aber zur Menge der Modifikatoren gehörten, die im Zusammenhang mit dem Adjektiv 'sicher' eingeschätzt wurden (Ausnahme: 'völlig').
5. Kategoriale Einstufungen (numerisch) und Magnitude-Einschätzungen (Linien und Zahlen) der subjektiven Wichtigkeit der 13 Berufswerte in Hinblick auf deren Realisierung im eigenen Beruf, um so ein weiteres, für die Berechnung individueller Urteilsfunktionen notwendiges Konstrukt einführen zu können.

Zu diesem Basisprogramm kamen schließlich noch zwei weitere Aufgaben, die eine axiomatische Überprüfung der Skalenqualität (vgl. hierzu ORTH 1982; WEGENER 1983; ORTH & WEGENER 1983) ermöglichen sollten:

6. Paar-Differenzen-Schätzungen, bei denen die Befragten die Größe der Wichtigkeitsdifferenz zwischen je zwei Itempaaren auf einer 20-stufigen Skala (1='sehr geringe Differenz', 20='sehr große Differenz') einstufen sollten, um auf diese Weise die zur Beurteilung des Meßniveaus der kategorialen Wichtigkeitsskalen notwendigen Zusatzinformationen zu gewinnen.
7. Paar-Verhältnisschätzungen, bei denen die Befragten in Bezug auf jedes Itempaar gebeten wurden, anzugeben, welches sie für wichtiger halten und anschließend zu sagen, um wieviel mal wichtiger sie das wichtigere Item halten. Diese Aufgabe diente dazu, die notwendigen Zusatzinformationen zur Beurteilung des Meßniveaus der Magnitude-Skalen zu gewinnen.

1.3 Das realisierte Forschungsprogramm

Pläne für internationale Forschungszusammenarbeit, ja selbst bereits getroffene Absprachen werden häufig von organisatorischen Sachzwängen eingeholt, die das Vorhaben stützen und die Bedeutung der dann noch möglichen Ergebnisse eindämmen. Für das vorliegende Projekt trifft dies leider auch zu. Es erwies sich nämlich als nicht durchführbar, daß der amerikanische Partner das Programm, über das man sich schriftlich bereits verständigt hatte, auch realisieren konnte. Auf einer gemeinsamen Arbeitssitzung beider Projektgruppen im März 1981 im NORC konnte für den amerikanischen Beitrag lediglich ein Minimalprogramm durchgesetzt werden, das aus den folgenden Aufgaben bestand:

- + Einstufungen aller 13 Berufswerte auf der numerischen Version der Wichtigkeitsskala;
- + Bi-modale Magnitude-Einschätzungen aller 13 Berufswerte.

Der Grund für diese Reduktion des Forschungsprogramms war in erster Linie der, daß NORC den zunächst mit ZUMA vereinbarten Plan, die Stichproben der Methodenstudie als von den Stichproben des ALLBUS bzw. GSS unabhängige Quotenstichproben zu realisieren, aufgab und die Stichprobe des amerikanischen Teils der Methodenstudie durch Zufallsauswahl aus der GSS-Stichprobe bestimmte. Die Befragten dieser Stichprobe sollten die Aufgaben der Methodenstudie als Appendix zu dem Fragenprogramm des GSS vorgelegt bekommen. Ein Fragenprogramm des oben geschilderten Umfangs hätte in diesem Fall das Gesamtinstrument von der zu erwartenden Interviewdauer her unanwendbar gemacht.

Dieses - relativ späte - Ausscheren aus der konzeptionellen Gemeinsamkeit, das auch die Interviewerschulung betraf (vgl. Abschnitt 2.1), hat für ein bedeutendes Forschungsinstitut wie das NORC durchaus eine nachvollziehbare Eigenlogik. Routine und Kontinuität müssen für ein solches Institut hohe Orientierungswerte darstellen, um für einen Auftraggeber Erwartbarkeit zu signalisieren. Die Gegenrechnung, daß nämlich die Bereitschaft, mit Neuem zu experimentieren dabei zweitrangig wird, hat bewußt nachgeordnete Priorität. Vor diesem Hintergrund darf es allerdings als Erfolg gewertet werden, daß das Team des GSS die Magnitude-Methodologie übernahm. In der bi-modalen Form und mit dem von ZUMA entwickelten Format wurde das Verfahren in dieser Untersuchung zum ersten Mal als Bestandteil einer nationalen Erhebung in den USA eingesetzt.

Was den deutschen Teil der Methodenstudie angeht, so stellte das Vorgehen von NORC für die ZUMA-Projektgruppe keinen Hinderungsgrund dar, die Realisierung des vollständigen Forschungsprogramms anzustreben. Das heißt, während die erhebungsmäßig parallelen Teile der Doppelstudie sich nach der amerikanischen Entscheidung nur noch auf die Kategorial- und Magnitudemessung sowie auf einen Teil des demographischen Hintergrundmaterials bezog, enthielt die ZUMA-Studie alle in Abschnitt 1.2 beschriebenen Aufgaben. Dies hatte jedoch zur Konsequenz, daß die Ergebnisse des deutschen Teils der Untersuchung nicht oder nur sehr eingeschränkt zur Beurteilung der Vergleichbarkeitsproblematik herangezogen werden konnten. Sie stellen vielmehr von dieser Problematik weitgehend unabhängig Beiträge dar.

2. PRAKTISCHE DURCHFÜHRUNG

2.1 Das Meßinstrument

Die in Abschnitt 1.2 beschriebenen Aufgaben führten zum Entwurf eines Erhebungsinstruments, das im deutschen Teil der Studie vollständig realisiert werden konnte. Dabei konnte - insbesondere, was die Magnitude-Methodologie angeht - in großem Umfang an Ergebnisse der ZUMA-Forschungen zur Methodenentwicklung angeknüpft werden, die sich in verschiedenen inhaltlichen Bereichen bewährt haben (vgl. WEGENER 1978, 1980, 1982; BECK et al. 1979; GRUNERT 1980, 1983a, 1983b).

Das entworfene Instrument bestand aus

- a. einem Fragebogen, der von den Interviewern auszufüllen war und zugleich die Instruktionen für die Befragten enthielt. Außerdem enthielt er einen demographischen Kurzteil, der neben einem Ausschnitt aus der ZUMA-Standarddemographie die Fragen zu politischen Einstellungen, subjektiver Schichteinstufung und 'Materialismus/Postmaterialismus' enthielt.
- b. einem Antwortheft, das der Befragte selbst durcharbeiten mußte und das neben zwei Übungsphasen für die Aufgaben der Magnitudeskalierung, bei der die Befragten Kreisgrößen bzw. die Schwere von Delikten mit Linien und Zahlen einstufen mußten. Die Formate wurden bereits in langjährigen Untersuchungen erprobt (vgl. WEGENER 1980).

Bei den kategorialen Einstufungen der Wichtigkeit wurden die Items mit Hilfe eines immer wieder neu gemischten Kärtchenspiels dargeboten. Die im Antwortheft gestellten unkonventionelleren Aufgaben zur Magnitude-Skalierung der Berufswerte, zu den Paar-Differenzen-Schätzungen und zu den Paar-Verhältnisschätzungen sind in den Abbildungen 1a-c beispielhaft dargestellt.

Abb. 1a: Format der Magnitude-Aufgabe zur Skalierung
der Berufswerte

Für wie wichtig halten Sie persönlich die folgenden
Merkmale für die berufliche Arbeit und den Beruf?

Hohes Einkommen

VERGLEICHSLINIE

Aufgaben, die viel Verantwortungsbewußtsein erfordern

ANTWORTLINIE

Gute Aufstiegsmöglichkeiten

ANTWORTLINIE

Viel Kontakt zu anderen Menschen

ANTWORTLINIE

Abb. 1b: Format der Aufgabe zur Differenzschätzung

Wie groß ist die **D i f f e r e n z** zwischen den beiden Berufsmerkmalen in Hinsicht auf die Wichtigkeit, die diese für Sie persönlich haben?

* GIBT MIR DAS GEFÜHL, ETWAS SINNVOLLES ZU TUN

* SICHERE BERUFSSTELLUNG

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
sehr geringe Differenz	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	sehr große Differenz

Wie groß ist die **D i f f e r e n z** zwischen den beiden Berufsmerkmalen in Hinsicht auf die Wichtigkeit, die diese für Sie persönlich haben?

* EIN BERUF, DER ANERKANNT UND GEACHTET WIRD

* HOHES EINKOMMEN

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
sehr geringe Differenz	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	sehr große Differenz

Abb. 1c: Format der Aufgabe zur Verhältnisschätzung

A. Welches der beiden folgenden Berufsmerkmale halten Sie persönlich für wichtiger?

B. Für wievielfach wichtiger halten Sie persönlich das angekreuzte Merkmal im Verhältnis zum anderen?

* SICHERE BERUFSSTELLUNG ☐

* GUTE AUFSTIEGSMÖGLICHKEITEN ☐

(beide gleich wichtig) ☐

(bitte Zahl
eintragen)

A. Welches der beiden folgenden Berufsmerkmale halten Sie persönlich für wichtiger?

B. Für wievielfach wichtiger halten Sie persönlich das angekreuzte Merkmal im Verhältnis zum anderen?

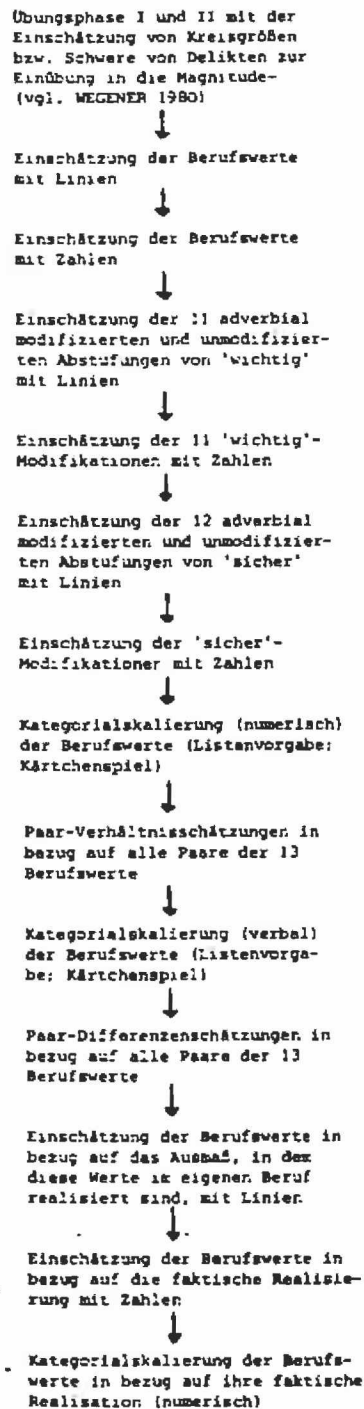
* AUFGABEN, DIE VIEL VERANTWORTUNGSBEWUSSTSEIN ERFORDERN ☐

* GIBT MIR DAS GEFÜHL, ETWAS SINNVOLLES ZU TUN ☐

(beide gleich wichtig) ☐

(bitte Zahl
eintragen)

Abb. 2: Zeitlicher Ablauf des Interviews



Im Falle der bi-modalen Magnitude-Skalierung der Berufswerte wurde als Standardreiz das Item 'Hohes Einkommen' festgelegt. Im Fall der lediglich im deutschen Teil der Untersuchung erfolgten Magnitude-Skalierung der adverbial modifizierten Adjektive 'sicher' und 'wichtig' waren die Standardreize 'einigermaßen sicher' bzw. 'einigermaßen wichtig'. Die Auswahl der Standards erfolgte in Übereinstimmung mit der im Bereich der Magnitude-Skalierung üblichen Strategie, als Standard solche Reize zu wählen, die im Median- bzw. Mittelwertsbereich der Antwortverteilung liegen. Die hierzu notwendigen Daten wurden im Fall der Berufswerte aus dem ALLBUS 1980 entnommen, im Fall der adverbial modifizierten Adjektive der Untersuchung von ROHRMANN (1978). Der exakte zeitliche Ablauf des Interviews ist Abbildung 2 zu entnehmen.

Die endgültige Festlegung der im amerikanischen und deutschen Teil der Methodenstudie zu applizierenden Meßinstrumente erfolgte - nachdem das entsprechende Erhebungsmaterial von ZUMA entworfen worden war und der NORC-Feldabteilung in Übersetzung vorlag - im Dezember 1981 im Rahmen einer Fragebogen- und Pretestkonferenz in der New Yorker Abteilung des Instituts unter Teilnahme eines ZUMA-Mitarbeiters. Die Festlegung geschah unter Berücksichtigung der offenbar besonderen Bedürfnisse amerikanischer Interviewer, die (soweit sie am Pretest teilgenommen hatten) auch zu Wort kamen. Im Unterschied zum Vorgehen in der deutschen Teilstudie war das Instrument aber auf eine schriftliche Schulung der Interviewer ausgelegt, weil ein nationaler, räumlich weit auseinandergezogener Interviewerstab zum Einsatz kommen sollte. Allerdings konnten auch in dieser Hinsicht ZUMA-Erfahrungen mit ähnlichen Untersuchungen weitergegeben werden. Neben dem Antwortheft für die Befragten und Interviewerinstruktionen bestand das Erhebungsmaterial auch aus einem 'Übungsheft für Interviewer' nach deutschem Vorbild (vgl. WEGENER 1980).

2.2 Stichproben

2.2.1 ZUMA-Studie

Nach Fertigstellung der umfangreichen Erhebungsinstrumente, Pretestungen zur Zeitabschätzung und mündlicher wie schriftlicher Interviewerschulung (29.7.1981) begann im August 1981 die vierwöchige Feldphase. Eingesetzt waren 13 Interviewer des ZUMA-Interviewerstabs. Die Stichprobe war als von der Stichprobe des ALLBUS

1982 unabhängige Quotenstichprobe ausgelegt. Sie wurde im Einzugsbereich von ZUMA realisiert (Mannheim, Heidelberg und Umgebung) und umfaßte 104 Befragte.

ZUMA entschied sich für eine Quotenstichprobe, um leere Zellenbesetzungen sowie Unterbesetzungen der Zellen zu vermeiden. Daß diese Gefahr bei einer Zufallsstichprobe dieser Größenordnung durchaus besteht, zeigt die Realisation der NORC-Stichprobe. Vergleiche zwischen der ZUMA-Stichprobe und der Stichprobe des ALLBUS 1982 sowohl auf der Ebene der Mittelwertsvergleiche (vgl. WEGENER 1983a) als auch auf der Konstruktebene (vgl. S.62, Tab. 20) ergaben eine recht hohe Ähnlichkeit der Verhältnisse der ZUMA-Stichprobe mit der NORC-Stichprobe.

Angesichts der kleinen Fallzahl wurde von einer ungefähren Gleichbesetzung der Zellen (bezüglich drei Alters- und Bildungsstufen sowie Geschlecht) ausgegangen. Eine leichte Überbesetzung der mittleren Alterskategorie war allerdings wünschenswert. Außerdem waren nur solche Befragte in die Stichprobe aufzunehmen, die gegenwärtig oder doch zumindestens zu einem früheren Zeitpunkt berufstätig waren, weil dies die Voraussetzung für die Beantwortung der Frage nach der 'faktischen Realisation' im eigenen Beruf war (vgl. Abschnitt 1.2, Aufgabe 5). Diese Forderung führte zu einer natürlichen Begrenzung der jüngsten Alterskohorten nach unten.

Der obere Teil von Tabelle 2 beschreibt die realisierten Ergebnisse. Die Definition der Bildungsstufen ist Tabelle 3 zu entnehmen.

Die Nichtberücksichtigung der über 65-jährigen ist nicht auf eine bewußte Beschränkung der Quotierung zurückzuführen, sondern ist die Folge einer mangelnden Kommunikation mit dem NORC und des Sachverhalts, daß die Interviewer keine Befragten über 65 aufsuchten. Auf der anderen Seite steht aufgrund vieler Untersuchungen fest, daß ältere Menschen zwar häufiger verweigern, aber in der Skalierungsqualität keine bedeutsamen Unterschiede aufweisen.

Tab. 2: Aufgliederung der Stichproben nach Alter, Geschlecht, Schicht und Bildung

		<u>MÄNNLICH</u>			<u>WEIBLICH</u>			
<u>BILDUNG</u>		niedrig	mittel	hoch	niedrig	mittel	hoch	TOTAL
<u>ALTER</u>								
<u>ZUMA</u>	18-30	7	5	5	5	5	5	32
	31-45	6	9	8	5	8	5	41
	46-65	5	5	5	6	5	5	31
	66-90	-	-	-	-	-	-	--
		18	19	18	16	18	15	104
<u>NORC</u>	18-30	1	7	1	1	9	6	25
	31-45	3	3	5	1	15	8	35
	46-65	4	7	2	3	9	5	30
	66-90	3	2	1	5	3	2	16
		11	19	9	10	36	21	106

2.2.2 NORC-Studie

Die amerikanische Erhebung wurde im Frühjahr 1982 als Teil der GSS-Hauptstudie durchgeführt. Die Auswahl der ca. 100 Befragten erfolgte als Auswahl eines 'random subsamples' der GSS-Gesamtstichprobe. Einzelheiten der Ziehungsprozedur wurden von NORC nicht weitergegeben. Die Zufallsauswahl führte zu 17% (nicht näher spezifizierter) Ausfälle im Magnitude-Addendum. Die gezogene NORC-Teilstichprobe enthielt 106 Fälle. Eine Aufteilung dieser Fälle nach den Quotierungsmerkmalen der deutschen Studie ergibt die im unteren Teil von Tabelle 2 beschriebenen Verteilungen.

Die Zuordnung der Bildungsstufen in den beiden Stichproben der Methodenstudie erfolgte in der Weise, daß den Schulabschlüssen der deutschen Stichprobe in der in Tabelle 3 dargestellten Weise Altersstufen in der amerikanischen Stichprobe zugeordnet wurden. Die Strategie, vergleichbare Bildungskategorien über die Zuordnung von Altersstufen zu Schulabschlüssen zu schaffen, wurde bereits in anderen Studien praktiziert (vgl. BARNES, KAASE et al. 1979).

Tab. 3: Definition des Bildungsgrades

	<u>ZUMA</u>	<u>NORC</u>
Niedrig	Volksschulabschluß (oder abgebrochen)	1 - 10 Jahre
Mittel	Mittlere Reife Fachhochschulreife	11 - 13 Jahre
Hoch	Abitur	14 - 20 Jahre

Eklatante Oberbesetzungen (z.B. Frauen in der Kategorie 31-45 Jahren mit mittlerem Schulniveau) fallen ebenso auf wie eklatante Unterbesetzungen: In acht Zellen finden sich nur zwei und weniger Befragte. Darüberhinaus weist die NORC-Stichprobe mehr ältere Befragte auf - eine Konsequenz der deutschen Quotenstichprobe.

2.2.3 Vergleichende Betrachtung der Verteilung einiger zentraler demographischer Variablen

Tabelle 4 vergleicht die Verteilungen der Befragten in beiden Stichproben der Methodenstudie in Hinsicht auf Alter, Geschlecht und Bildung untereinander und in bezug auf die entsprechenden Ergebnisse des ALLBUS 1982 und des GSS 1982.

(Spalten 1-4). Diese Ergebnisse werden hier zur Orientierung mitgeteilt. In ihnen spiegeln sich vor allem die unterschiedlichen Ziehungsprozeduren. Dabei muß nochmals auf den bedauerlichen Umstand hingewiesen werden, daß der amerikanische Partner ursprünglich zugesagt hatte, ebenso wie bei der ZUMA-Studie eine Quotenstichprobe zugrunde zu legen. In diesem Fall wären Diskontinuitäten in den Zellenbesetzungen der NORC-Stichprobe vermieden worden.

Tab. 4: Vergleich wichtiger Stichproben und Teilstichproben bezüglich Alter, Geschlecht und Bildung

	1	2	3	4	5	6	7	8
<u>ALTER</u>	ZUMA	NORC	ALL82	GSS82	NORC-Y	NORC NOMAG	ZUMA SIGN	NORC SIGN
18-30	30.8	23.6	21.1	19.5	26.8	16.7	47.1	25.0
31-45	39.4	33.0	28.7	25.6	28.6	16.7	35.3	37.5
46-65	29.8	28.3	31.2	27.5	33.9	61.1	17.6	18.5
66-90	--	15.1	19.1	17.3	10.7	5.6	--	18.5
N	104	106	2994	1494	56	18	17	16
<u>GESCHLECHT</u>								
MÄNNLICH	52.9	37.1	44.8	42.4	23.2	27.8	47.1	37.5
WEIBLICH	47.1	62.9	55.2	57.6	76.8	72.2	52.9	62.5
N	104	106	2994	1606	56	18	17	16
<u>BILDUNG</u>								
NIEDRIG	32.7	19.8	63.5	23.4	17.9	38.9	23.5	6.3
MITTEL	35.6	51.9	25.2	49.7	51.8	44.4	41.2	31.2
HOCH	31.7	28.3	11.3	26.9	30.3	16.7	35.3	62.5
N	104	106	2953	1494	56	18	17	16
					(52.8%)	(17%)	(16.3%)	(18.2%)

In Spalte 5 kommt eine unbeabsichtigte Besonderheit des amerikanischen Datensatzes zum Ausdruck: Die Analyse dieses Datensatzes, der bei ZUMA erstellt wurde, ergab nämlich sofort, daß nur 56 Befragte der 106 Befragten der NORC-Stichprobe (das sind 52.8%) die Magnitude-Skalierung und die kategorischen Einschätzungen aller 13 Berufswerte vorgenommen hatten. Der verbleibende Teil hatte stattdessen die 5 ursprünglichen amerikanischen Items ihrer Rangordnung nach bewertet. Dieser Umstand erklärt sich daraus, daß die amerikanische Teilstudie als Bestandteil der GSS-Erhebung auch dem Split-Verfahren in bezug auf den Gesamtsurvey unterworfen war. Es war nicht darauf geachtet worden, daß nur solche Befragten den Magnitude-Zusatz erhielten, die im Hauptteil des Interviews auch die 13-Item-Version der Berufswerte-Frage (Y-Version der Frage) vorgelegt bekamen. Diese Unachtsamkeit als Folge einer mechanischen Stichprobenziehung führte zu einer Halbierung der Fälle der amerikanischen Stichprobe, die für einen Skalenvergleich zwischen Magnitude- und Kategorialmessung zur Verfügung standen. Aus Spalte 5 in Tabelle 5 (NORC-Y) wird ersichtlich, daß diese eingeschränkte Gruppe verbleibender Befragter sich in etwa wie die Hauptpopulation in Bezug auf Alter, Geschlecht und Bildung verteilt ist. Ein Vergleich der Faktorenstrukturen der Magnitude-Messungen zwischen den Befragten mit der Y-Version der Berufswerte-Frage und allen Befragten ergab im übrigen eine bemerkenswerte Ähnlichkeit der Faktorstrukturen.

Spalte 6 von Tabelle 5 (NORC NOMAG) schlüsselt die Ausfälle der amerikanischen Stichprobe in Bezug auf den Magnitude-Zusatz auf. Während sich diese Ausfälle ungefähr proportional auf die verschiedenen Befragtengruppen verteilen (mit Ausnahme der Alterskategorie 46-65 Jahre und der unteren Bildungsgruppe), ist hervorzuheben, daß die Gesamtausfallrate von 17% knapp doppelt so hoch ist wie jene, die bei vergleichbaren repräsentativen Untersuchungen in der Bundesrepublik auftraten (vgl. u.a. WEGENER 1980). Wir wissen, daß Ausfälle und die Weigerung, das Magnitude-Verfahren anzuwenden, in Umfragen in erster Linie durch das Interviewverhalten bedingt sind. Berücksichtigt man, daß die Methode hier erstmals in den USA und unter den Bedingungen der dort üblichen Interviewerschulung (schriftlich) angewandt wurde, kann die Auswahlrate keinesfalls als unerwartet hoch eingestuft werden.

Die beiden letzten Spalten von Tabelle 5 (ZUMA SIGN und NORC SIGN) geben die Verteilungen und Häufigkeiten der 'signifikanten' Magnitude-Skalierungen wieder. Dabei ist eine bi-modale Magnitude-Messung signifikant, wenn der

Exponent der Zwischenmodalitäten-Funktion (ICMM-Funktion), die die auf beiden Modalitäten gewonnenen Einschätzungen zueinander in Beziehung setzt, signifikant von einem theoretisch erwarteten Wert abweicht (nach oben oder nach unten). Obwohl solche signifikanten Skalen für die Auswertung nicht schlechthin unbrauchbar sind, zeigen sie doch an, daß die impliziten Verhältnissbildungen in den beiden benutzten Modalitäten (Linien und Zahlen) Inkonsistenzen aufweisen. Insgesamt traten in beiden Untersuchungen insgesamt 16.3% bzw. 18.2% solcher signifikanten Abweichungen auf. Wichtiger als diese Beträge, die etwas höher als erwartet liegen (auf dem .05-Niveau), ist jedoch, daß sich die signifikanten Messungen nicht auffällig bei irgendwelchen Alter-, Geschlechts- oder Bildungsgruppen häufen. Die 'Güte' der Skalierungen ist nicht von diesen Merkmalen der Befragten abhängig. Vielmehr zeigen andere Untersuchungen, daß auch hier der Interviewer-Einfluß den Effekt bestimmt.

3. ERGEBNISSE DER SEMANTISCHEN UNTERSUCHUNGEN

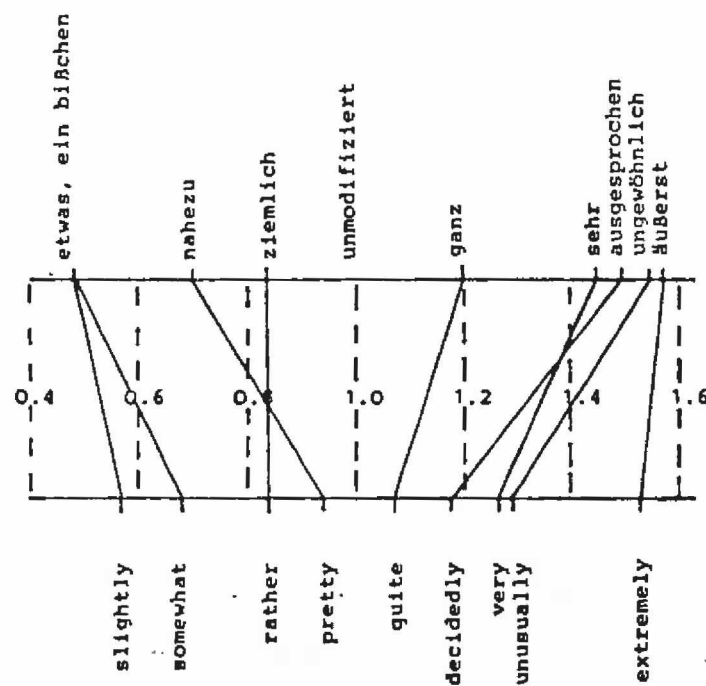
3.1 Einleitende Bemerkungen

Wie bereits in Abschnitt 1.2 ausgeführt wurde, sah das ursprüngliche Forschungsprogramm der Methodenstudie die Untersuchung der Frage vor, inwiefern sich internationale Unterschiede in den Wirkungen adverbialer Modifikatoren auf die Vergleichbarkeit der Messungen auswirken. Diese Frage wurde einerseits in Hinblick auf die Wirkung adverbialer Modifikatoren im Rahmen der Itemformulierungen und andererseits in Hinblick auf die Wirkung adverbialer Modifikatoren im Rahmen von Antwortvorgaben verbaler Kategorienskalen gestellt. Die untersuchten Adjektive und ihre Modifikatoren wurden bereits vorgestellt (vgl. S.5).

Daß das Verständnis adverbial modifizierter Adjektive und die Wirkung adverbialer Modifikatoren in internationalen Untersuchungen kontrolliert zu werden verdient, konnte KRISTOF (1966) bei einem Vergleich zwischen deutschen und englischen Adverbien zeigen. KRISTOF arbeitete mit Wörterbuchübersetzungen der von CLIFF (1960)

in einer internationalen Studie verwendeten Adverbien. In Abbildung 3 sind die modifizierenden Wirkungsstärken der betrachteten deutschen und amerikanischen Adverbien auf zwei Axen markiert. Die entsprechenden Wörterbuch-Übersetzungen sind durch Linien verbunden. Man sieht sofort, daß die deutschen Adverbien in keiner Weise ihren amerikanischen Übersetzungen entsprechen. Die Modifikationsbereiche sind ebenfalls verschieden. Die deutschen Adverbien schließen einen größeren Bereich ein. Selbst die ordinalen Relationen bleiben nicht überall erhalten. Allerdings steht zu erwarten, daß die Werte der Modifikatorstärken mit den betrachteten Adjektiven wechseln. KRISTOF und CLIFF betrachteten das Adjektiv 'günstig' ('favourable').

Abb. 3: Modifikatorstärken deutscher Adverbien und ihrer englischen Übersetzungen (aus KRISTOF 1966)



Da in der vorliegenden Studie die Adjektive 'sicher' im Kontext des Items 'Sichere Berufsstellung' und 'wichtig' betrachtet wurden, wären hier möglicherweise andere Ergebnisse zu erwarten gewesen.

Nachdem sich herausgestellt hatte, daß die Untersuchung der Wirkungsstärken adverbialer Modifikatoren im internationalen Rahmen nicht mehr zu realisieren war und die hierzu erforderlichen Messungen nur noch im deutschen Teil der Studie durchführbar waren (vgl. WEGENER, FAULBAUM & MAAG 1982a, 1982b), konzentrierten sich die Bemühungen auf

- den Nachweis der Einsatzfähigkeit der Magnitude-Skalierung zur Bestimmung der Modifikatorstärken. Die Verwendung der Magnitude-Skalierung zur subjektiven Einschätzung der Wirkungsstärken adverbialer Modifikatoren würde - im Gegensatz zu den üblicherweise verwendeten aggregierten Skalen - den Vorteil einer Individualskalierung auf höherem Meßniveau bieten. Da die Magnitude-Skalierung aber in einer expliziten psychophysischen Theorie wurzelt, wäre zuerst nachzuweisen, daß die Einschätzung der subjektiven Intensitäten adverbial modifizierter Adjektive im Einklang mit der psychophysischen Theorie steht.
- die Untersuchung der Abhängigkeiten der Modifikatorstärken von den wichtigsten demographischen Merkmalen wie 'Alter', 'Geschlecht', 'Bildung' und 'Sozialer Status', um so herauszufinden, ob die bei Befragungen naiv unterstellte Annahme der Invarianz der Modifikatorstärken berechtigt ist oder nicht.

3.2 Modelle zur Vorhersage und Erklärung der Wirkungsstärken adverbialer Modifikatoren

Zur Erklärung und Vorhersage der Wirkungsstärke adverbialer Modifikatoren wurde in der Literatur (HOWE 1962; LILLY 1968a, 1968b; KRISTOF 1966) bisher ausschließlich das von CLIFF (1959) erstmals vorgeschlagenen lineare Modell herangezogen. Dieses Modell hat die folgende Gestalt:

$$(1) \quad Q_{ij}(MA) = w_{M_j} Q_i(A) + K,$$

wobei

$Q_{ij}(MA)$: Skalenwert des i -ten Adjektivs in Kombination mit dem j -ten Modifikator;

w_{M_j} : Multiplikationskoeffizient für den j -ten Modifikator;

$Q_i(A)$: der psychologische Skalenwert des i -ten Adjektivs;

K : Differenz zwischen dem beliebigen Nullpunkt der Skalenwerte und ihrem psychologischen Nullpunkt.

Die empirische Bestimmung der Parameter von (1) erfolgte in den genannten Untersuchungen auf der Basis der Methode der sukzessiven Intervalle, also einer Methode, die keine Bestimmung interindividueller Unterschiede erlaubt, da die Skalen in diesem Fall stets für alle Individuen gemeinsam ermittelt werden. Eine zweite Implikation des Untersuchungsansatzes liegt darin, daß sich die Skalen der modifizierten und unmodifizierten Adjektive auf alle Adjektive, die in die Untersuchung eingehen, zugleich beziehen. Schließlich geht in die gemeinsame Skalenkonstruktion der modifizierten und unmodifizierten Adjektive unüberprüfbar die Unterstellung ein, daß die Wirkung der Modifikatoren der einzelnen Adjektive identisch ist.

Vor dem Hintergrund dieser Kritik wurde in der Methodenstudie ein Ansatz vorgeschlagen, der die funktionalen Beziehungen zwischen modifizierten und unmodifizierten Adjektiven kontextspezifisch und bezogen auf Individualdaten zu bestimmen versucht. Das bedeutet, daß die Skalierungen auf Individualniveau vorgenommen werden müssen und daß die Modifikatorwirkungen stets im Kontext jeweils spezifischer Adjektive betrachtet werden. Der vorgeschlagene Ansatz legt als Skalierungsverfahren die Methode der bi-modalen Magnitude-Skalierung zugrunde. Dabei ergibt sich unter der Annahme, daß sich die Wirkung adverbialer Modifikatoren auf unmodifizierte Adjektive als Beziehung zwischen unterschiedlichen, intensiven Reaktionsmodalitäten interpretieren läßt, für die die Befunde des cross-modality matching Paradigmas gelten, für jedes Individuum L die folgende Potenzbeziehung:

$$(2) Q_{\ell}(MA) = kQ_{\ell}(A)^{w_M}$$

bzw. als:

$$(3) \log Q_{\ell}(MA) = w_M \log Q_{\ell}(A) + \log k.$$

3.3 Zusammenfassende Darstellung der Befunde

Die wesentlichen Ergebnisse der semantischen Analysen im deutschen Teil der Methodenstudie lassen sich wie folgt zusammenfassen:

1. Die Befunde in bezug auf die indirekte cross-modality matching Funktion und in bezug auf die Interskalenrelationen (zwischen Kategorial- und Magnitude-Skalen) bestätigen, daß die Eigenschaften der Magnitude-Skalen in Übereinstimmung mit den Resultaten der sensorischen und sozialen Psychophysik stehen und daß sich die Inzensitätswahrnehmungen von 'sicher' und 'wichtig' wie 'prothetische' Kontinuen verhalten. Im Falle der Zwischenmodalitätenbeziehungen ergeben sich mit hoher Anpassungsgüte Potenzfunktionen mit 1 als Exponenten. In Tabelle 5 sind die Ergebnisse im einzelnen wiedergegeben. In der Tabelle finden sich sowohl die aggregierten Werte, die sich bei der geometrischen Mittelung von Zahlen- und Linienreaktionen über alle 104 Personen ergeben, als auch die mittleren Parameter für die Anpassungen auf Individualniveau (es sind die die Mediankorrelationen eingetragen). Die Spalte 'Linearität' enthält die F-Werte der Linearitätstests. Die unter der Kategorie 'Präferenz' aufgeführten Zeilen enthalten die Parameter der Intermodalitätenfunktionen für die Zahlen- und Linienreaktionen der Wichtigkeitseinstufungen der Berufswerte. Diese Werte entsprechen denen von Tabelle 15.

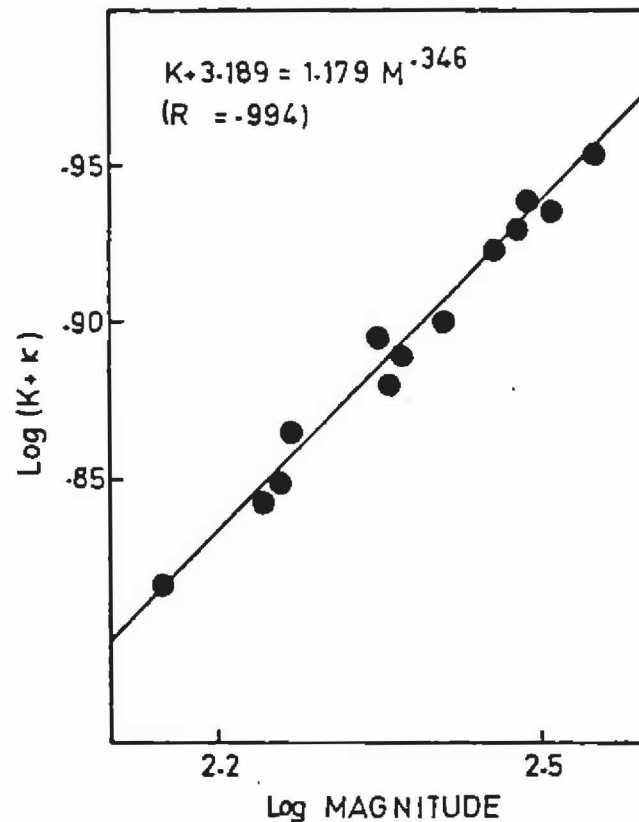
Auch im Fall der Interskalenrelationen zwischen den Magnitude- und kategorischen Messungen findet man bestätigt, daß sich die Skalen ganz im Sinne der häufig belegten Befunde der psychophysischen Forschung verhalten.

Tab. 5: Parameter der ICMM-Funktion für die Adjektive sicher und wichtig

	EXPONENT	KORRELATION	LINEARITÄT
<u>SICHER</u>			
AGGR.	1.070	.998	2110.0
INDIV.	1.084	.942	235.9
(S.D.)	(.333)	(.138)	(575.1)
<u>WICHTIG</u>			
AGGR.	.984	.999	4154.0
INDIV.	1.034	.928	176.7
(S.D.)	(.392)	(.131)	(349.5)
<u>PRAEFERENZ</u>			
AGGR.	.973	.979	258.8
INDIV.	1.059	.905	148.6
(S.D.)	(.556)	(.169)	(386.8)

In dem Maße nämlich, in dem die Magnitude-Skalen logarithmische Intervallskalen und in dem die kategorischen Skalen in bezug auf identische Stimulusserien einfache Intervallskalen darstellen, ergibt sich als Funktion zwischen beiden eine additive Potenzfunktion der Form $K_i + \kappa = aM_i^\alpha$ mit K_i als kategorische und M_i als Magnitude-Skalenwerte (WEGENER & KIRSCHNER 1981; WEGENER 1982a). Abbildung 4 zeigt die Anpassung dieser Funktionen mit den geschätzten Parametern für die aggregierten Magnitude- und Kategorialskalen der Merkmalspräferenzen. Mit einer Korrelation von .994 kann die Gültigkeit der Interskalenrelation als hinreichend gesichert gelten. Auf Individualniveau lassen sich entsprechende Funktionen anpassen; die mittlere Korrelation für alle individuellen Kurvenanpassungen ist .892 (S.D.= .155).

Abb. 4: Interskalenrelation zwischen den Kategorial- (K) und Magnitude-Skalen (M)



Die in Abbildung 4 dargestellte Interskalenfunktion stellt im übrigen einen interessanten Sonderfall dar: Die Magnitude-Skala ist eine 'indirekte' Magnitude-Skala; sie ergibt sich, wenn man die verbalen kategorischen Antwortvorgaben (sehr wichtig, ziemlich wichtig etc.), die gewählt wurden, durch die entsprechenden Magnitude-Werte der getrennt vorgenommenen wichtig-Skalierung ersetzt.

2. Die regressionsanalytische Schätzung des in (3) auftretenden w_m -Parameters führt zu einer intuitiv einsichtigen Rangfolge der Modifikatorwirkungen. (vgl. Tabelle 6).

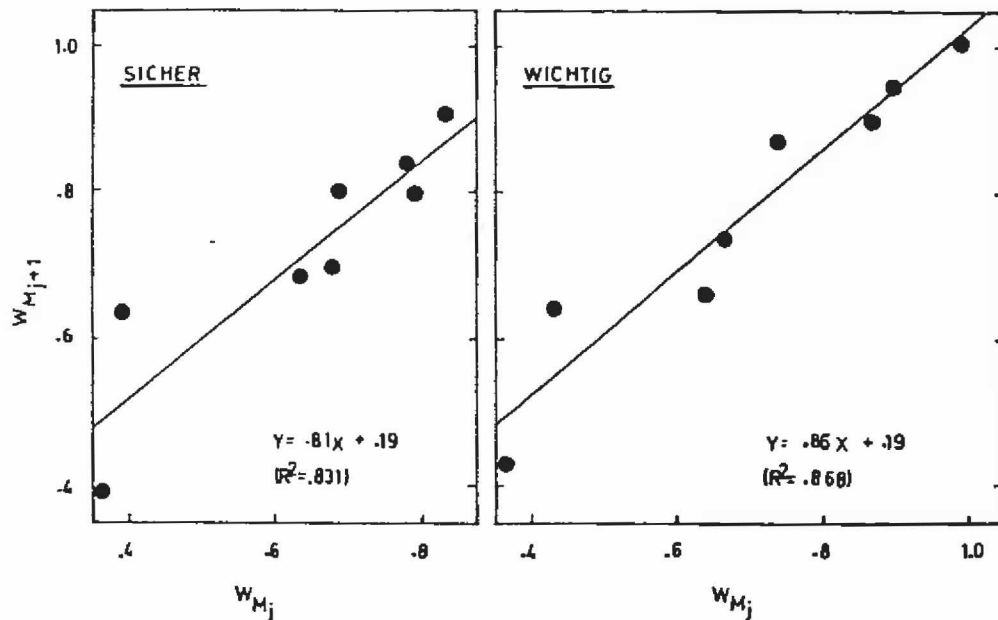
Tab. 6: Exponenten (w_m) für die Modifikatoren und Korrelationskoeffizienten $^m(R)$; vgl. Gleichungen (2) und (3)

	<u>SICHER</u>		<u>WICHTIG</u>	
	w_M	R	w_M	R
außerordentlich	.904	.823	1.063	.882
sehr	.837	.846	.944	.887
ziemlich	.785	.857	.900	.874
überwiegend	.801	.855	.871	.853
halbwegs	.693	.849	.739	.817
einigermaßen	.681	.823	.665	.789
teilweise	.641	.789	.640	.753
etwas	.388	.672	.429	.649
wenig	.361	.618	.361	.495
Mittelwerte	.677	.792	.735	.778

Insgesamt ergibt sich für die adverbialen Modifikatoren eine stärkere Wirkung bei wichtig als bei sicher, was die Vermutung bestätigt, daß die Wirkung adverbialer Modifikatoren nicht invariant ist gegenüber einer Veränderung des Adjektivs.

3. Die Wirkung eines Modifikators verhält sich zur Wirkung des nächst stärkeren Modifikators proportional, und zwar unabhängig vom gewählten Adjektiv: Die Verhältnisse der Exponenten für wichtig und sicher sind paarweise in etwa identisch, während sich die Wirkungen einzelner adverbialer Modifikatoren in bezug auf die zu modifizierenden Dimensionen unterscheiden (vgl. Abbildung 5
4. Schließlich finden sich systematische Unterschiede in der Modifikatorwirkung zwischen den nach Sozialstatus, Alter und Geschlecht quotierten Befragten-
gruppen, sodaß von einer Invarianz des Verständnisses von adverbialen Modifikatoren nicht ausgegangen werden kann. Eine genauere Darstellung dieser Befunde findet sich in WEGENER, FAULBAUM & MAAG (1982a).
5. Es gibt Hinweise für eine Wechselwirkung der Wirkungsstärken stimulus- und reaktionsseitiger Modifikatoren.

Abb. 5: Regression von w_{M_j} auf $w_{M_{j+1}}$



4. ERGEBNISSE DES VERGLEICHS ZWISCHEN NUMERISCHEN UND VERBALEN KATEGORIALSKALEN

4.1 Einleitende Bemerkungen

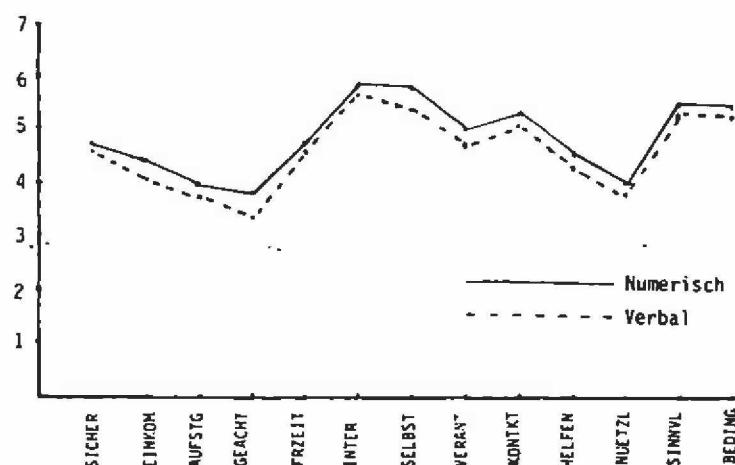
Nachdem die Untersuchung der Auswirkungen unterbleiben mußte, die eine Ersetzung der im ALLBUS bzw. GSS verwendeten 7-stufigen Kategorienskala mit numerischen Antwortvorgaben durch eine 7-stufige Skala mit den in Abschnitt 1.2 beschriebenen verbalen Antwortvorgaben auf die internationale Vergleichbarkeit hat, mußte sich der Vergleich der beiden Skalenformen ausschließlich auf die deutsche Stichprobe beschränken. Dabei interessierten uns vor allem die folgenden Fragestellungen:

- Welche Berufsmerkmale können im Rahmen eines geeigneten Meßmodells als Indikatoren welcher Konstrukte aufgefaßt werden, bzw. welche latenten Konstrukte werden durch die Wichtigkeitseinstufungen welcher Berufsmerkmale erfaßt?
- Wie 'rein' werden die Konstrukte durch die entsprechenden Merkmale erfaßt, bzw. gibt es zwischen den verbal und numerisch erfaßten Merkmalen ins Auge fallende Reliabilitätsunterschiede?
- Inwieweit messen die von uns betrachteten numerischen und verbalen Skalen identische Konstrukte (Problem der Konstruktinvarianz)?

4.2 Betrachtung der Einzelitems

Eine Betrachtung der Einzelitems hinsichtlich ihrer Mittelwerte ergibt zunächst, daß bei einer nahezu gleichen Rangordnung die Mittelwerte der verbalen Skala für alle Items durchweg niedriger liegen als die der numerischen Skala (vgl. die Mittelwertsprofile in Abbildung 6).

Abb. 6: Mittelwertsprofile der numerischen und verbalen Kategorienskala



Allerdings sind nur fünf dieser Unterschiede signifikant (WILCOXON-Test für abhängige Stichproben). Das Gleiche, was für das Verhältnis der Mittelwerte zueinander gilt, gilt auch für das Verhältnis der Mediane zueinander. Insgesamt gesehen zeigt sich also, daß die mit Hilfe der verbalen Skala ermittelten Meßwerte im Mittel durch die Werte der numerischen Skala unterschätzt werden.

Zur Orientierung sei mitgeteilt, daß alle Items eine - allerdings nur mäßige - negative Verteilungsschiefe aufweisen, wobei die Verteilungsschiefen der verbalen Skala für die allermeisten Items geringer sind. Eine Ausnahme bildet lediglich das Item 'Interessante Tätigkeit'. Im Fall von 'Hohes Einkommen' und 'Ein Beruf, der anerkannt und geachtet wird' kehren sich die Verteilungsschiefen um (numerisch: negative Schiefe; verbal: positive Schiefe). Allerdings ist diese Umkehrung wohl aufgrund der jeweils niedrigen Werte zu vernachlässigen.

Tab. 7: Ergebnisse der Regressionsanalysen

	<u>b</u>	<u>R</u>
Sichere Berufsstellung	0.707	0.744
Hohes Einkommen	0.772	0.798
Gute Aufstiegsmöglichkeiten	0.712	0.717
Ein Beruf, der anerkannt und geachtet wird	0.532	0.580
Ein Beruf, der einem viel Freizeit läßt	0.784	0.784
Interessante Tätigkeit	0.525	0.451
Eine Tätigkeit, bei der man selbstständig arbeiten kann	0.710	0.627
Aufgaben, die viel Verantwortungsbewußtsein erfordern	0.707	0.677
Viel Kontakte zu anderen Menschen	0.747	0.719
Ein Beruf, bei dem man anderen helfen kann	0.742	0.808
Ein Beruf, der für die Gesellschaft nützlich ist	0.800	0.832
Gibt mir das Gefühl, etwas Sinnvolles zu tun	0.719	0.652
Sichere und gesunde Arbeitsbedingungen	0.700	0.682

Aufschluß darüber, inwieweit sich die Skalenwerte der verbalen Skalen denen der numerischen linear vorhersagen lassen, geben die Ergebnisse der für die einzelnen Items getrennt durchgeführten Regressionsanalysen (vgl. Tabelle 7), bei denen die verbalen Versionen der Skalen als Funktion der numerischen ausgedrückt wurden. Berechnet man den Mittelwert der in Tabelle 7 aufgeführten R's, so ergibt sich ein Wert von .704. Insgesamt zeigt sich, daß der lineare Zusammenhang nicht für alle Items gleich gut ist. Für einige Berufswerte ist der Determinationskoeffizient nicht größer als .2 oder .3 (bei 'Interessante Tätigkeit' z.B. $R^2 = .203$). Als mittlerer Determinationskoeffizient ergibt sich ein R^2 von .497. Dieses Ergebnis zeigt, daß wir nicht ohne weiteres davon ausgehen können, daß kategoriale Skalen mit verbalen Antwortvorgaben und numerisch benannte Skalen identische, lediglich linear transformierte Ergebnisse liefern.

4.3 Identifikation zugrundeliegender Konstrukte

4.3.1 Ergebnisse der exploratorischen Faktorenanalysen

Tabelle 8 vermittelt einen Überblick über die Ergebnisse der exploratorischen Faktorenanalysen, wobei auch die Ergebnisse für die numerische Skala des ALLBUS 1982 dargestellt sind. Die Faktorenanalysen wurden nach dem Hauptkomponentenverfahren ohne iterierte Kommunalitäten gerechnet. Die Ladungen sind VARIMAX-rotiert. Die vier höchsten Ladungen auf einem Faktor sind jeweils unterstrichen.

Wie man sieht, ergaben die Analysen für alle drei Skalen eine relativ klare dreidimensionale orthogonale Struktur, die der schon im ALLBUS 1982 gefundenen Struktur weitgehend entspricht (vgl. SCHMIDT 1982). Die einzelnen Faktoren wurden dort als 'Extrinsische Orientierung' (EXT), 'Intrinsische Orientierung' (INT) und 'Soziale Orientierung' (SOZ) bezeichnet.

Der größte Varianzanteil wird in allen drei Fällen durch den Faktor EXT aufgeklärt, der geringste durch den Faktor SOZ. Während sich die durch die einzelnen Faktoren aufgeklärten Varianzanteile bei der numerischen und der verbalen Skala der Methodenstudie in etwa gleicher Größenordnung bewegen, fällt beim ALLBUS die weitaus größere Dominanz des ersten Faktors auf (etwa doppelt so hohe Varianzaufklärung).

Tab. 8: Ergebnisse der exploratorischen Faktorenanalysen

Variablen	AJJAK 1982 NUMERISCH			SKALLENVERGLEICH NUMERISCH			SKALLENVERGLEICH VERBAL		
	EXT	INT	SKZ	EXT	INT	SKZ	EXT	INT	SKZ
Sichere Berufsaufstellung	<u>0.714</u>	0.348	0.114	<u>0.726</u>	-0.033	0.270	<u>0.714</u>	-0.162	<u>0.311</u>
Hohes Einkommen	<u>0.835</u>	0.085	0.070	<u>0.864</u>	0.011	-0.179	<u>0.839</u>	-0.086	-0.221
Gute Aufstiegsmöglichkeiten	<u>0.807</u>	0.249	0.174	<u>0.847</u>	0.211	0.113	<u>0.873</u>	0.010	0.018
Ein Beruf, der anerkannt und geachtet wird	<u>0.697</u>	0.295	0.280	<u>0.694</u>	0.051	0.280	<u>0.666</u>	-0.049	0.271
Ein Beruf, der einem viel Freizeit läßt	0.517	0.352	0.113	0.006	0.442	-0.409	-0.100	0.190	-0.559
Interessante Tätigkeit	0.371	<u>0.771</u>	0.184	-0.073	<u>0.743</u>	0.041	-0.080	<u>0.762</u>	-0.253
Eine Tätigkeit, bei der man selbständig arbeiten kann	0.241	<u>0.831</u>	0.177	0.036	<u>0.821</u>	0.060	-0.040	<u>0.778</u>	-0.242
Aufgaben, die viel Verantwortungsbewußtsein erfordern	0.295	<u>0.669</u>	0.340	0.221	<u>0.708</u>	0.223	0.031	<u>0.669</u>	0.124
Viel Kontakte zu anderen Menschen	0.087	0.355	<u>0.637</u>	0.140	<u>0.537</u>	0.385	-0.086	<u>0.631</u>	0.302
Ein Beruf, bei dem man anderen helfen kann	0.148	0.156	<u>0.877</u>	0.199	0.009	<u>0.840</u>	-0.038	0.381	<u>0.746</u>
Ein Beruf, der für die Gesellschaft nützlich ist	0.207	0.175	<u>0.831</u>	0.342	-0.059	<u>0.738</u>	0.044	-0.102	<u>0.848</u>
Gibt mir das Gefühl, etwas Sinnvolles zu tun	0.176	<u>0.601</u>	<u>0.517</u>	-0.025	0.349	<u>0.707</u>	-0.269	0.570	<u>0.353</u>
Sichere und gesunde Arbeitsbedingungen	0.401	0.574	0.297	0.022	0.218	<u>0.419</u>	0.076	0.076	0.258

Die Struktur der in Tabelle 8 dargestellten Ladungsmatrizen wird auch durch andere Verfahren der explorativen Faktorenanalyse im wesentlichen bestätigt. Im Falle von mit EFAP II (vgl. JÖRESKOG & SÖRBOM 1978) gerechneten Analysen bestätigen die Größenordnungen der TUCKER-LEWIS-Koeffizienten (verbal: 0.901; numerisch: 0.889), daß das dreifaktorielle Modell sowohl im Falle der numerischen als auch im Falle der verbalen Version der Skala die beobachtete Varianz-Kovarianz-Matrix sehr gut erklärt.

Betrachtet man die Durchschnittsmengen der vier höchsten Ladungen über alle drei Skalen, so ergeben sich für die einzelnen Konstrukte die folgenden Kernitems:

Extrinsische Orientierung

Sichere Berufsstellung
Hohes Einkommen
Gute Aufstiegsmöglichkeiten
Ein Beruf, der anerkannt und geachtet wird

Intrinsische Orientierung

Interessante Tätigkeit
Eine Tätigkeit, bei der man anderen helfen kann
Aufgaben, die viel Verantwortungsbewußtsein erfordern

Soziale Orientierung

Ein Beruf, bei dem man anderen helfen kann
Ein Beruf, der für die Gesellschaft nützlich ist

Der Faktor 'Extrinsische Orientierung' scheint also im wesentlichen den Karriereaspekt zu kennzeichnen, während der Faktor 'Intrinsische Orientierung' im wesentlichen den Aspekt der persönlichen Erfüllung kennzeichnet. Der Faktor 'Soziale Orientierung' betrifft das persönliche Engagement für die Gesellschaft und den Mitmenschen.

Neben den über die Analysen hinweg invarianten Kernitems gibt es in den Fällen der 'Intrinsischen Orientierung' und der 'Sozialen Orientierung' weitere Items, die diese Faktoren in unterschiedlichem Ausmaß charakterisieren. So gehört im Falle der ALLBUS-Untersuchung das Item 'Gibt mir das Gefühl, et-

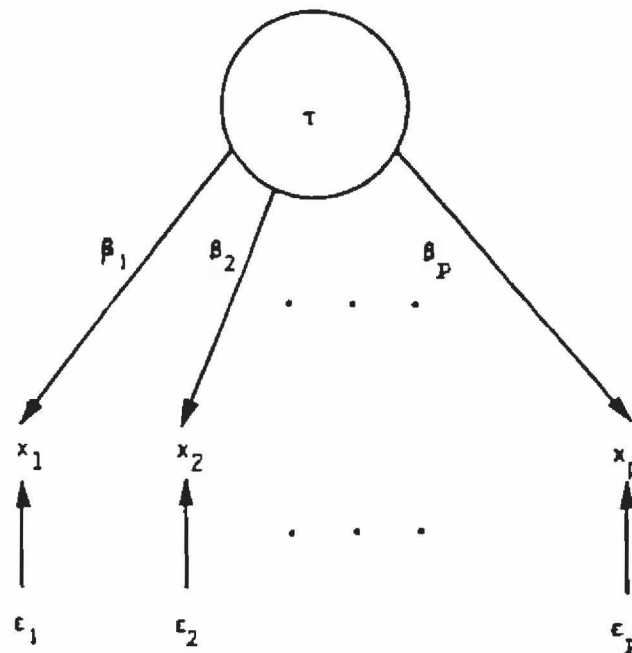
was Sinnvolles zu tun" zu den vier höchsten Ladungen, während dies bei den Skalen der Methodenstudie nicht der Fall ist. Dort gehört hingegen das Item "Viel Kontakt zu anderen Menschen" zu den vier höchsten Ladungen. Ähnliches ergibt sich, wenn wir den sozialen Faktor betrachten, wo das Item "Viel Kontakte zu anderen Menschen" im Fall des ALLBUS zu den vier höchsten Ladungen zählt, nicht aber im Falle der beiden Skalen der Methodenstudie. Schließlich gibt es gravierende Unterschiede zwischen dem ALLBUS und den beiden Skalen der Methodenstudie auch bezug auf das Item "Ein Beruf, der einem viel Freizeit läßt". Im Gegensatz zu ALLBUS kennzeichnet dieses Item den sozialen Faktor bei den Skalen der Methodenstudie stark negativ. Beim ALLBUS läßt dieses Item relativ hoch auf dem extrinsischen Faktor. Die geschilderten Unterschiede sind möglicherweise auf eine unterschiedliche Besetzung der Stichproben zurückzuführen.

4.3.2 Identifikation von Indikatoren

Nachdem - einer weitverbreiteten Praxis folgend - mittels exploratorischer Faktorenanalysen ein Überblick über die den Items zugrundeliegenden möglichen Konstrukte hergestellt wurde, erhebt sich die Frage, welche Items als Indikatoren für diese Konstrukte angesehen werden können. Als notwendige Bedingung für die Zulassung eines Items als 'Indikator' für ein Konstrukt sehen wir dabei an, daß das Item und das betreffende Konstrukt positiv miteinander kovariieren. Betrachtet man mehrere Items, so besteht eine weitere notwendige Bedingung darin, daß die Items ein kongenerisches Meßmodell bilden. Dies tun sie dann, wenn ihre wahren Werte mit 1.0 miteinander korrelieren (vgl. JÖRESKOG 1971, 1978; ALWIN & JACKSON 1980; CARMINES & McIVER 1981; FAULBAUM 1983; NASH-PARKER 1983). In diesem Fall nämlich lassen sich die Itemmeßwerte als Funktion der Werte eines einzigen latenten Konstrukts auffassen (vgl. Abbildung 7).

Bei der Identifikation der Indikatoren sind wir zunächst von der Frage ausgegangen, ob es für die im letzten Abschnitt erwähnten Konstrukte der Extrinsischen Orientierung, der Intrinsischen Orientierung und der Sozialen Orientierung kongenerische Meßmodelle gibt, die der numerischen Skala des ALLBUS, der numerischen Skala der Methodenstudie und der verbalen Skala der Methodenstudie gemeinsam sind.

Abb. 7: Schema eines kongenerischen Meßmodells



Das methodische Vorgehen bestand dabei darin, daß wir für jedes Konstrukt solche Items in Betracht gezogen haben, die bei mindestens einer der drei Skalen zu den vier höchsten positiven Ladungen zählen und eine Ladung größer als 0.6 besitzen. Nach diesem Auswahlkriterium ließen sich unter den einzelnen Konstrukten folgende Items subsumieren:

Extrinsische Orientierung:

Sichere Berufstellung
Hohes Einkommen
Gute Aufstiegsmöglichkeiten
Ein Beruf, der anerkannt und geachtet wird

Intrinsische Orientierung:

Interessante Tätigkeit
Tätigkeit, bei der man selbständig arbeiten kann
Aufgaben, die viel Verantwortungsbewußtsein erfordern
Viel Kontakte zu anderen Menschen
Gibt mir das Gefühl, etwas Sinnvolles zu tun

Soziale Orientierung:

Viel Kontakte zu anderen Menschen
Ein Beruf, bei dem man anderen helfen kann
Ein Beruf, der für die Gesellschaft nützlich ist
Gibt mir das Gefühl, etwas Sinnvolles zu tun

Die Überprüfung der Frage, inwieweit es sich bei diesen Itemkonstellationen um kongenerische Meßmodelle handelt, ergab die in Tabelle 9 dargestellten Ergebnisse. Unter Zugrundelegung einer Entscheidungsregel, bei der nur solche Meßmodelle akzeptiert werden, die einen GFI (Goodness-of-fit-Index; vgl. JÖRESKOG & SÖRBOM 1981: I.40) größer 0.950 und einen RMR (root mean square residual; vgl. JÖRESKOG & SÖRBOM 1981: I.41) kleiner 0.100 besitzen, müssen wir akzeptieren, daß die Annahme eines kongenerischen Meßmodells nur im Falle des ALLBUS uneingeschränkt erfüllt ist. Für die übrigen Fälle sind die Bedingungen für das Vorliegen eines kongenerischen Meßmodells nur für das Konstrukt Extrinsische Orientierung erfüllt. Bei der verbalen Skala zeigte sich im Falle der Sozialen Orientierung, daß die Evaluation des Modells zu einer 'improper solution' in Form negativer Fehlervarianzen führte. Mit derartigen Lösungen ist im Falle relativ kleiner Stichproben gehäuft zu rechnen, wie Simulationsstudien ergaben (vgl. BOOMSMA 1982).

In den Fällen, wo die Annahme eines kongenerischen Meßmodells erfüllt war, sahen wir uns berechtigt, für die einzelnen Items nach dem in FAULBAUM (1983) näher beschriebenen Verfahren die Reliabilitäten zu bestimmen. Die Tabellen 10, 11 und 12 enthalten neben den Reliabilitäten (ρ_i) die Ladungen (β_i) und die Fehlervarianzen (θ_i^2).

Tab. 9: Anpassungsindizes für die kongenerischen Modelle

	ALLBUS 1982 NUMERISCH	SKALENVERGLEICH NUMERISCH	SKALENVERGLEICH VERBAL
<u>EXT</u>			
χ^2	5.160 (df=2; P=0.076)	0.860 (df=2; P=0.650)	0.480 (df=2; P=0.786)
χ^2/df	2.580	0.430	0.240
GF1	0.998	0.995	0.997
RMR	0.020	0.041	0.031
<u>INT</u>			
χ^2	78.320 (df=5; P=0.000)	10.290 (df=5; P=0.067)	10.490 (df=5; P=0.062)
χ^2/df	15.660	2.060	2.100
GF1	0.975	0.951	0.850
RMR	0.057	0.117	0.139
<u>SOZ</u>			
χ^2	57.580 (df=2; P=0.000)	6.260 (df=2; P=0.004)	-
χ^2/df	28.790	3.130	-
GF1	0.981	0.968	-
RMR	0.061	0.135	-

Wie Tabelle 10a sowie 11 und 12 zeigen, erreicht im Falle des Konstrukts Extrinsische Orientierung das Item 'Gute Aufstiegsmöglichkeiten' sowohl bei der numerisch abgestuften Skala des ALLBUS als auch bei den numerisch- bzw. verbal abgestuften Skalen der Methodenstudie konsistent die höchste Einzelreliabilität, sodaß man davon ausgehen kann, daß das Konstrukt durch dieses Item am reinsten gemessen wird.

Tab. 10: Kongenerische Meßmodelle /ALLBUS 1982

a. EXTRINSISCH

	B_i	θ_i^2	ρ_i
Sichere Berufsstellung	1.188	1.075	0.567
Hohes Einkommen	1.016	1.032	0.502
Gute Aufstiegs- möglichkeiten	1.417	0.778	0.721
Ein Beruf, der aner- kannt und geachtet wird	1.223	1.209	0.553

b. INTRINSISCH

	B_i	θ_i^2	ρ_i
Interessante Tätigkeit	1.107	0.643	0.656
Tätigkeit, bei der man selbständig arbeiten kann	1.154	0.563	0.703
Aufgaben, die viel Verant- wortungsbewußtsein erfordern	1.157	0.866	0.607
Viel Kontakte zu anderen Menschen	0.827	1.588	0.366
Gibt mir das Gefühl, etwas Sinnvolles zu tun	0.935	0.952	0.479

c. SOZIAL

	B_i	θ_i^2	ρ_i
Viel Kontakte zu anderen Menschen	0.922	1.412	0.376
Ein Beruf, bei dem man anderen helfen kann	1.331	0.769	0.697
Ein Beruf, der für die Gesellschaft nützlich ist	1.273	0.850	0.656
Gibt mir das Gefühl, etwas Sinnvolles zu tun	0.896	1.022	0.440

Tab. 11: Kongenerisches Meßmodell/Skalenvergleich Numerisch

EXTRINSISCH

	β_i	θ_i^2	ρ_i
Sichere Berufsstellung	1.053	1.600	0.409
Hohes Einkommen	1.259	1.332	0.543
Gute Aufstiegs- möglichkeiten	1.276	0.468	0.776
Ein Beruf, der aner- kannt und geachtet wird	1.093	1.750	0.406

Tab. 12: Kongenerisches Meßmodell/Skalenvergleich Verbal

EXTRINSISCH

	β_i	θ_i^2	ρ_i
Sichere Berufsstellung	0.980	1.508	0.389
Hohes Einkommen	1.254	1.116	0.585
Gute Aufstiegs- möglichkeiten	1.229	0.542	0.736
Ein Beruf, der aner- kannt und geachtet wird	0.880	1.688	0.521

Im Falle der numerischen Skala des ALLBUS ergaben sich beim Konstrukt Intrinsische Orientierung die niedrigsten Reliabilitäten bei 'Viel Kontakte zu anderen Menschen' bzw. 'Gibt mir das Gefühl, etwas Sinnvolles zu tun'. Im Fall des Konstrukts Soziale Orientierung ergaben sich die niedrigsten Reliabilitäten ebenfalls bei diesen Items.

Nachdem der Versuch, für alle drei Skalen (Allbus/Numerisch, Skalenvergleich/Numerisch und Skalenvergleich/Verbal) in bezug auf die Konstrukte Intrinsische Orientierung und Soziale Orientierung gemeinsame kongenerische Meßmodelle zu bilden, gescheitert war, wurde untersucht, ob es für die beiden Skalen der Methodenstudie (Numerisch, Verbal) gemeinsame Meßmodelle gibt. Unsere Bemühungen die sich auf ein in FAULBAUM (1984:48) näher erläutertes Vorgehen stützten, führten schließlich zu den in Tabelle 13 dargestellten gemeinsamen kongenerischen Subskalen. Im Falle der Sozialen Orientierung führte die Analyse trivialerweise zu einem gemeinsamen kongenerischen Meßmodell, da ein Modell mit drei Items gerade identifiziert ist und keine Freiheitsgrade mehr zur Verfügung stehen.

Tab. 13: Gemeinsame kongenerische Modelle der numerischen und verbalen Skala des Skalenvergleichs

a. INTRINSISCH

	<u>NUMERISCH</u>			<u>VERBAL</u>		
	β_i	θ_i^2	ρ_i	β_i	θ_i^2	ρ_i
Interessante Tätigkeit	0.527	0.721	0.278	0.776	0.808	0.427
Tätigkeit, bei der man selbständig arbeiten kann	0.965	0.450	0.674	1.004	0.752	0.573
Aufgaben, die viel Verantwortungsbewußtsein erfordern	1.124	0.492	0.720	0.955	1.007	0.475
Viel Kontakte zu anderen Menschen	0.906	1.925	0.299	0.828	2.322	0.228
$\chi^2 = 1.86$ (df=2; P=0.394)			$\chi^2 = 1.89$ (df=2; P=0.388)			
$\chi^2/df = 0.93$			$\chi^2/df = 0.95$			
GFI = 0.991			GFI = 0.991			
RMR = 0.034			RMR = 0.052			

b. SOZIAL

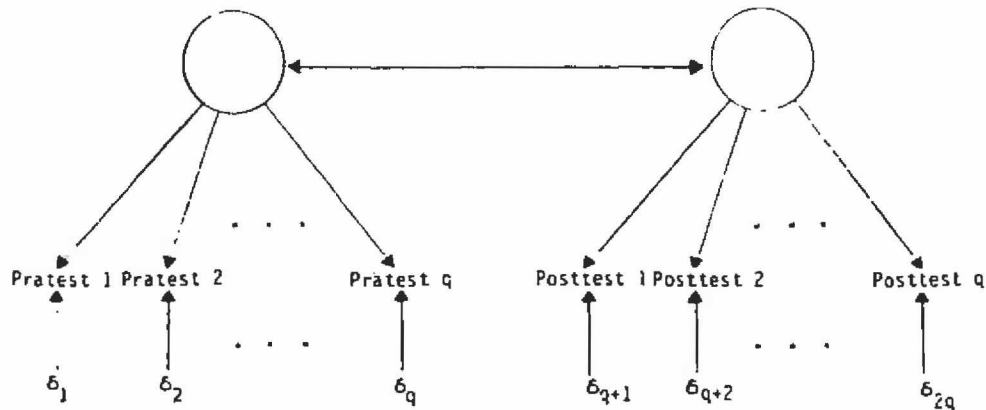
	<u>NUMERISCH</u>			<u>VERBAL</u>		
	β_i	θ_i^2	ρ_i	β_i	θ_i^2	ρ_i
Ein Beruf, bei dem man anderen helfen kann	1.558	0.762	0.761	1.536	0.252	0.903
Ein Beruf, der für die Gesellschaft nützlich ist	1.307	1.397	0.550	0.963	1.955	0.322
Gibt mir das Gefühl, etwas Sinnvolles zu tun	0.881	1.201	0.393	0.692	1.958	0.197

Eine vergleichende Betrachtung der Itemreliabilitäten bei den im den Tabellen 11, 12 und 13 beschriebenen kongenerischen Modellen zeigt, daß es durchaus substantiellere Unterschiede zwischen der numerischen und der verbalen Skala gibt, sodaß, wenn die Skalenpaare Identisches messen sollten, sie dies zumindest mit unterschiedlichen Itemreliabilitäten tun (Näheres vgl. FAULBAUM 1984:51).

4.4 Konstruktinvarianz numerischer und verbaler Skalen

Für die Behandlung der Frage, inwieweit die im letzten Abschnitt beschriebenen Paare numerischer und verbaler kongenerischer Meßmodelle Identisches messen, bieten sich jene Strukturgleichungsmodelle an, die JÖRESKOG & SÖRBOM (vgl. JÖRESKOG & SÖRBOM 1976; JÖRESKOG 1979) zur Analyse von Test-Retest-Situationen empfohlen haben. Diese Modelle setzen voraus, daß man einer Stichprobe von N Befragten zu zwei unterschiedlichen Gelegenheiten q kongenerische Tests appliziert hat. Zwischen beiden Gelegenheiten liegt die eigentliche Versuchsbedingung. Bezeichnen wir die q Tests in ihrer ersten Anwendung als Prätests und in ihrer zweiten Anwendung als Posttests, so erhalten wir - sofern wir an der Korrelation zwischen den Konstrukten interessiert sind - das in Abbildung 8 dargestellte Modell.

Abb. 8: Schema eines Modells mit q Prätests und q Posttests



In unserem Fall besteht die Versuchsbedingung in der Ersetzung der numerischen Kategorienbenennungen durch die entsprechenden verbalen Benennungen. Die beiden Gelegenheiten sind die beiden Zeitpunkte, an denen die entsprechenden numerischen und verbalen Items im Rahmen der Befragung eingestuft wurden. Insgesamt ergaben sich so drei Modelle (je eines für das Konstrukt der Extrinsischen Orientierung, der Intrinsischen Orientierung und der Sozialen Orientierung). Im Falle der Extrinsischen Orientierung z.B. gibt es vier kongenerische Prätests und vier kongenerische Posttests. Die kongenerischen Prätests sind die vier numerischen Items, die kongenerischen Posttests sind die vier verbalen Items. Das zum ersten Zeitpunkt durch die vier numerischen Items gemessene Konstrukt wollen wir als Extrinsische Orientierung - Numerisch (Abk.: EXT-N) und das durch die vier verbalen Items gemessene Konstrukt als Extrinsische Orientierung - Verbal (Abk.: EXT-V) bezeichnen. Entsprechend wie beim Konstrukt der Extrinsischen Orientierung gehen wir auch bei den übrigen Konstrukten vor. Die dort gewählten Kurzbezeichnungen sind INT-N bzw. INT-V und SOZ-N bzw. SOZ-V.

Unter der Voraussetzung, daß die entsprechenden kongenerischen numerischen und verbalen Skalen Identisches messen, würde man erwarten, daß die zugrundeliegenden Konstrukte EXT-N und EXT-V, INT-N und INT-V sowie SOZ-N und SOZ-V nahe 1.0 miteinander korrelieren.

Tab. 14: Anpassungsindizes

a. Extrinsische Orientierung

	χ^2	χ^2 / df	GFI	RMR	D-GFI
Modell- variante 1	139.36 df=19; P=0.000	7.33	0.773	0.092	-
Modell- variante 2	22.19 df=15; P=0.000	1.48	0.920	0.045	0.147

b. Intrinsische Orientierung

	χ^2	χ^2 / df	GFI	RMR	D-GFI
Modell- variante 1	101.18 df=19; P=0.000	5.325	0.821	0.101	-
Modell- variante 2	17.830 df=15; P=0.271	1.189	0.944	0.039	0.123

c. Soziale Orientierung

	χ^2	χ^2 / df	GFI	RMR	D-GFI
Modell- variante 1	111.40 df=8; P=0.000	13.925	0.755	0.118	-
Modell- variante 2	5.24 df=5; P=0.387	1.048	0.963	0.025	0.208

Wir haben versucht, die geschilderten Modelle in zwei verschiedenen Varianten anzupassen. Die erste Variante bestand in der Annahme, daß die Meßfehler der Items unkorreliert sind, die zweite Variante bestand in der Annahme, daß es, auf das Vorhandensein itemspezifischer Faktoren hindeutende, Korrelationen zwischen den Meßfehlern der entsprechenden Items der beiden Zeitpunkte gibt (also z.B. 'Sichere Berufsstellung', numerisch gemessen und 'Sichere Berufsstellung', verbal gemessen). Die Ergebnisse der Modellanpassungen für die beiden Modellvarianten sind in den Tabellen 14a-c dargestellt. Die dort enthaltenen Spalten D-GFI geben den Zuwachs des Goodness-of-fit-Indexes an, wenn wir statt des Modells ohne Meßfehlerkorrelationen das Modell mit Meßfehlerkorrelationen anpassen.

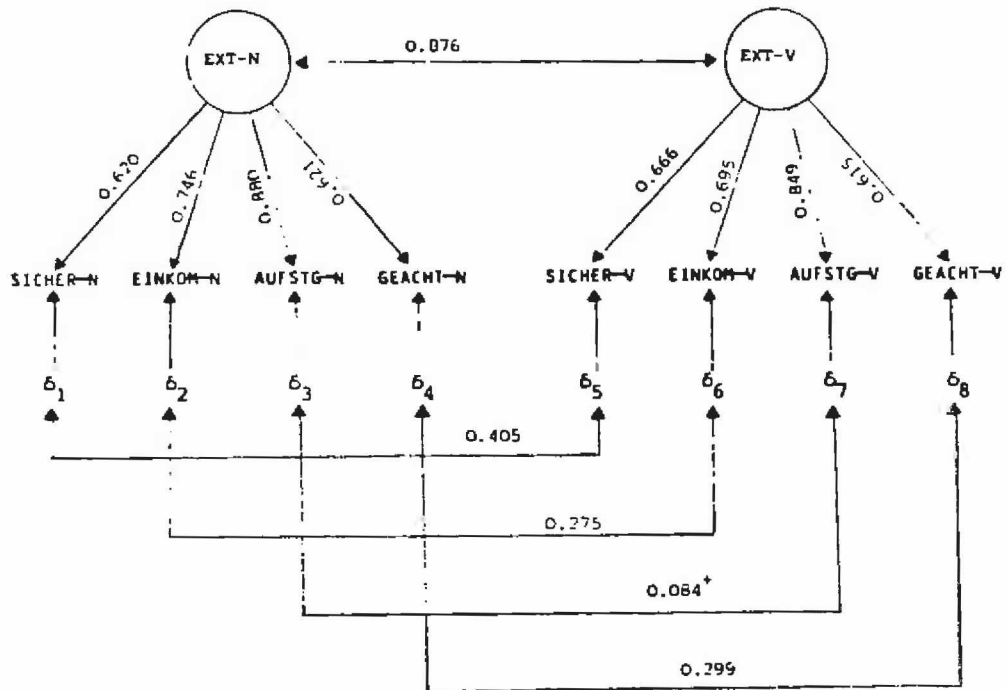
Die Tabellen 14a-c zeigen, daß das Modell mit der Annahme von Meßfehlerkorrelationen die Beobachtungsdaten am besten beschreibt, was auf Faktoren hindeutet, die keinen Beitrag zu den Korrelationen zwischen den beiden Zeitpunkten, sondern nur zu den Korrelationen zwischen den entsprechenden Items der numerischen und verbalen Skalen leisten (vgl. JÖRESKOG, 1979: 334).

In den Abbildungen 9a-c sind die Konstruktkorrelationen, die Faktorenladungen und die Fehlerkorrelationen eingetragen. Nicht signifikante Werte (T-Werte unter 2) sind dort mit '+' gekennzeichnet. Dargestellt sind die standardisierten Werte. Die Analysen basieren auf den Korrelationsmatrizen. Aus den in den Abbildungen verwendeten Abkürzungen lassen sich die vollständigen Itemformulierungen ohne Schwierigkeiten rekonstruieren.

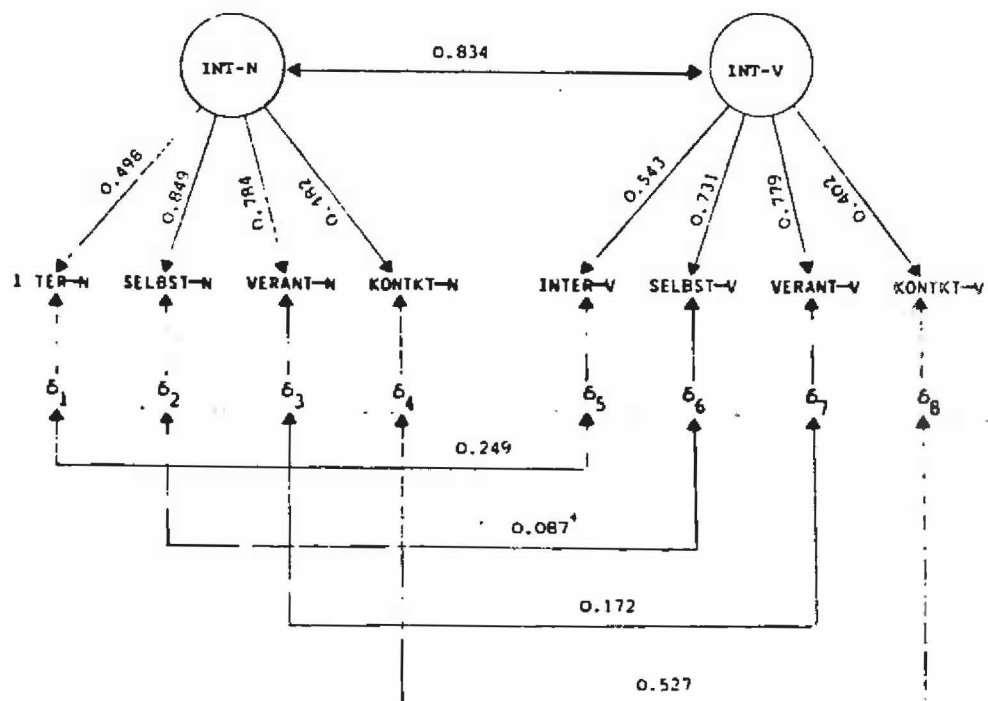
Es fällt auf, daß die Korrelationen zwischen den latenten Konstrukten zwar sehr hoch sind, daß sie aber im Fall der Extrinsischen Orientierung und der Intrinsischen Orientierung bei Einbeziehung der Meßfehlerkorrelationen in das Modell deutlich von 1.0 verschieden sind. Bei der Interpretation dieser Korrelationen muß man auch berücksichtigen, daß trotz der Tatsache, daß zwischen den numerischen und verbalen Wichtigkeitseinstufungen eine die Aufmerksamkeit der Befragten voll beanspruchende Aufgabe lag, das Vorliegen eines systematischen, auf die Korrelationen der wahren Werte durchschlagenden Erinnerungsfaktors nicht ausgeschlossen werden kann, so daß mit einer

Abb. 9: Modelle mit korrelierten Meßfehlern

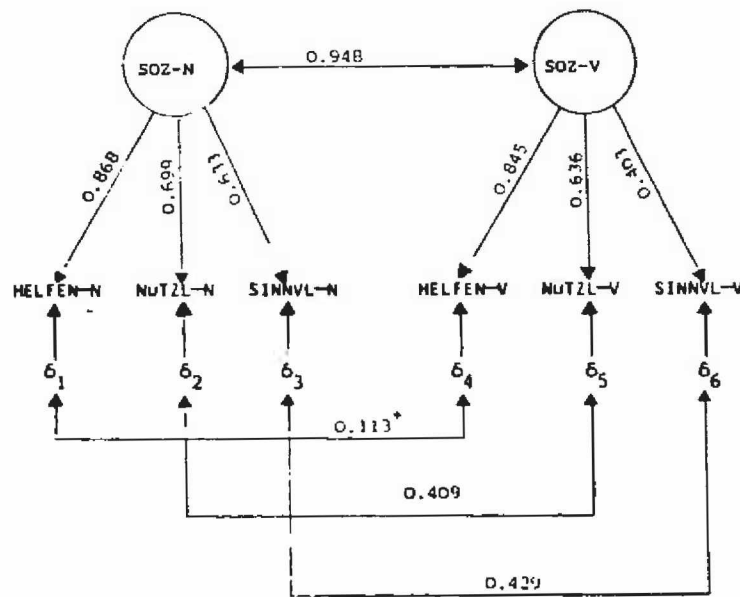
a. Extrinsische Orientierung



b. Intrinsische Orientierung



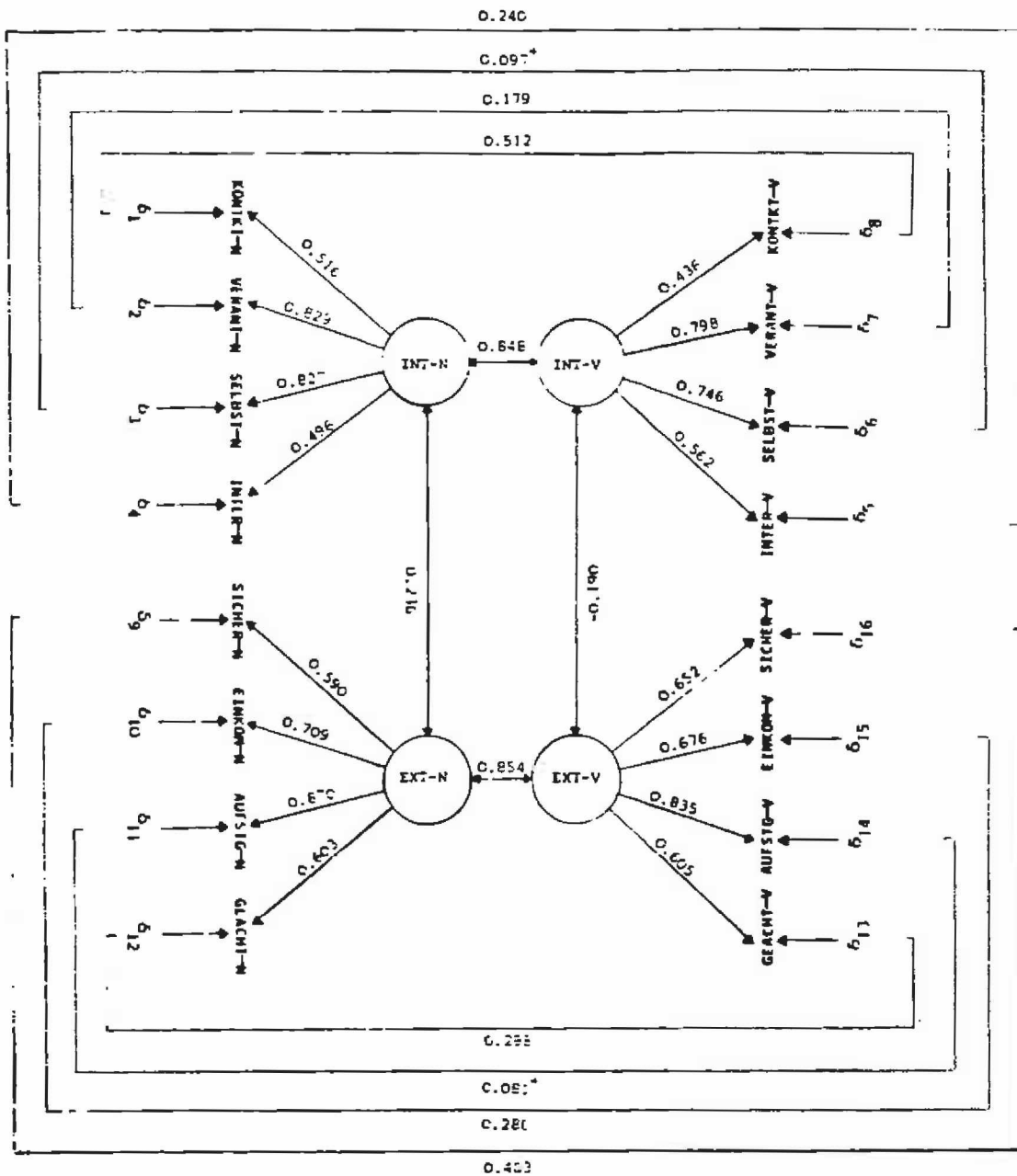
c. Soziale Orientierung



gewissen Überschätzung der wahren Korrelationen zu rechnen wäre. Aufgrund dieser Überlegungen über den Zusammenhang der gemessenen Konstrukte wird deutlich, daß beide Skalierungsformen hinsichtlich der gemessenen Inhalte nicht unreflektiert durcheinander substituiert werden können.

In dieser Bewertung wird man bestärkt, wenn man untersucht, wie sich die Ersetzung der numerischen Skala durch die von uns verwendete verbale Skala auf die korrelativen Beziehungen zwischen mehreren Konstrukten auswirkt. In Abbildung 10 ist ein Modell mit zwei Konstrukten (EXT und INT) zu jedem Zeitpunkt unter der Bedingung von Fehlerkorrelationen dargestellt. Wie man sieht, ergibt sich zwischen EXT-N und INT-N eine signifikant von 0 verschiedene positive Korrelation, während die Korrelation zwischen EXT-V und INT-V negativ ist; d.h. eine Ersetzung der numerischen Einstufungsform durch die verbale kann sogar zu einer Umkehrung der korrelativen Beziehungen zwischen den Konstrukten führen. Auch hier wurde der Analyse die Korrelationsmatrix zugrundegelegt.

Abb. 10: Modell mit vier Konstrukten
(EXT-N, INT-N, EXT-V, INT-V)



Anpassungsindizes:

$$\chi^2 = 131.60 \text{ (df=92; P=0.004)}$$

$$\chi^2/\text{df} = 1.43$$

$$\text{GFI} = 0.754$$

$$\text{RMR} = 0.076$$

5. DIE INTERNATIONALE VERGLEICHBARKEIT DER BERUFSWERTE-SKALA

5.1 Vorbemerkungen

Nachdem in den letzten beiden Kapiteln über die Ergebnisse derjenigen Problemstellungen berichtet wurde, die nur im deutschen Teil der Studie untersucht werden konnten, kommen in diesem Kapitel im wesentlichen jene Ergebnisse zur Darstellung, die sich auf die Erhebungen in beiden Nationen stützen und die sich auf die eigentliche Vergleichbarkeitsproblematik beziehen. Die Struktur der Darstellung folgt dabei im großen und ganzen der in den Abschnitten 1.1 und 1.2 beschriebenen Logik des Vergleichs von Einstellungsskalen. Nach einer Diskussion der Skalenqualitäten der verwendeten Kategorial- und Magnitudeskalen wenden wir uns der Identifikation der hinter den Einschätzungen der Berufswerte in beiden Nationen möglicherweise verborgenen latenten Konstrukte zu und versuchen zu ermitteln, durch welche Indikatoren sich diese Konstrukte messen lassen. Unter der Voraussetzung der Vergleichbarkeit der Einschätzungen der Berufswerte würde man in beiden Nationen das Vorliegen gleicher Meßmodelle erwarten. Das Kapitel wird abgeschlossen durch einen Abschnitt über einige diskriminative Skaleneigenschaften.

5.2 Die vergleichende Beurteilung der meßtheoretischen Qualitäten

5.2.1 Die Magnitude-Messungen

Was die Magnitude-Methodologie angeht, so besteht einer ihrer Vorteile darin, daß sie sich aus einer empirisch bewährten Theorie ableitet, der psychophysischen Theorie. Die Einzelheiten dieser Theorie müssen hier nicht näher ausgeführt werden (vgl. z.B. STEVENS 1975; unter strukturellen Gesichtspunkten WEGENER 1982b). Für unseren Zusammenhang wichtig ist aber, daß diese Theorie drei hervorgehobene Kriterien zur Verfügung stellt, um die Qualität von Magnitude-Einstufungen zu bewerten.

Das erste Kriterium ist die Zwischenmodalitäten-Korrelation (ICMM (Indirect cross-modality matching)-Korrelation). Bei bi-modaler Magnitude-Messung ist offensichtlich, daß erfolgreiche Messungen durch eine hohe Korrelation zwischen

den beiden benutzten Modalitäten auszeichnen. Das zweite Kriterium ist die Form der Funktion, die sich zwischen den Meßwerten der beiden Modalitäten ergeben muß. Die Magnitude-Theorie verlangt hier eine Potenzfunktion. Das dritte Kriterium schließlich bezieht sich auf den Exponenten dieser Potenzfunktion. Aufgrund unzähliger psychophysischer Experimente hat sich erwiesen, daß dieser Exponent für bestimmte Modalitäten typisch und stabil ist. In unserem Fall, in dem wir mit Linien- und Zahlen-Reaktionen messen, wird erwartet, daß der Exponent die Größe 1 aufweist.

Die Erfüllung dieser drei Kriterien ist (unter Zusatzannahmen) hinreichend für die Bestimmung der Skalenqualität von Magnitude-Messungen. Erst eine Überprüfung der Messungen auf axiomatischer Grundlage könnte allerdings auch die notwendigen Bedingungen dafür bereitstellen.

Die Ergebnisse der numerischen Überprüfung im Sinne der genannten drei Kriterien für die Einschätzungen der 13 Berufswerte sind in Tabelle 15 zusammengestellt. Man kann die Kriterien an die aggregierten und an die individuellen Skalen anlegen. Im oberen Teil der Tabelle finden sich die Kennwerte der aggregierten Skalen, und zwar für die ZUMA- und NORC-Populationen getrennt (Spalten 1 und 2). Mit EXPO sind die Exponenten der ICMM-Funktion bezeichnet, mit CORR die Zwischenmodalitäten-Korrelationen, mit INTC die entsprechenden Interzepte, wenn der Zusammenhang im doppelt-logarithmischen Koordinatensystem berechnet wird (Y: Zahlen; X: Linien). F-LIN drückt den Wert der F-Prüfstatistik für den Linearitätstest der log-log-Funktion aus (d.h. für den Test auf Vorliegen der Potenzfunktion) und N die Anzahl der Fälle, für die die Magnitude-Skalen der 13 Berufswerte vorlagen.

Die Werte zeigen, daß die Zwischenmodalitäten-Korrelationen für die aggregierten Skalen in beiden Teilpopulationen hoch sind. Die F-Werte weisen beide aggregierten ICMM-Funktionen als Potenzfunktionen aus, und die Exponenten dieser Funktionen sind nicht signifikant von 1 verschieden, dem Wert, der bei der Verwendung von Linien und Zahlen zu erwarten ist. Allerdings ist das diesbezügliche Ergebnis für die deutsche Population günstiger als für die amerikanische.

Aufschlußreicher ist der mittlere Teil von Tabelle 15, der die Resultate individueller Analysen und die Variabilitäten (in Klammern) wiedergibt. Wir sehen,

Tab. 15: Indizes der Intermodalitäten-Beziehungen

MAGSKALEN INDICES	1 ZUMA	2 NORC	3 NORC-Y	4 NORC SELECT	5 ZUMA SELECT
AGGR					
EXPO	.973	.801			
CORR	.979	.965			
INTC	4.320	.539			
F-LIN	258.8	149.1			
N	104	88			
INDIV					
EXPO	1.059 (.556)	1.131 (.908)	1.161 (1.114)	1.114 (.395)	1.111 (.574)
CORR	.834 (.905*) (.169)	.657* (.806*) (.376)	.631 (.795*) (.398)	.753 (.836*) (.198)	.827 (.905*) (.159)
INTC	7.381 14.672)	1.358 (3.583)	1.113 (2.629)	.341 (.401)	7.981 (16.327)
F-LIN	148.6 (386.8)	39.8* (66.58)	45.79 (82.49)	36.73 (56.73)	103.5 (207.7)
N	104	88	48	62 (58.5%)	82 (76.8%)
RANGE-MAG	8.95 (11.28)	7.44 (14.93)	7.16 (15.27)	8.99 (17.50)	9.03 (11.39)
RANGE-KAT	4.31 (1.19)	---	3.37* (1.59)	3.55 (1.82)	4.43 (1.16)

* Median

* $p \leq .001$ im Vergleich zu ZUMA

daß insbesondere die ICM-Korrelationen in der NORC-Stichprobe niedriger liegen als bei der ZUMA-Stichprobe. Ebenso finden sich im Mittel auch Abweichungen von der Form der Potenzfunktion der Intermodalitäten-Relationen (F-LIN). Diese Unterschiede gelten nicht nur für alle Befragten der NORC-Untersuchung, sondern auch für nur jene, die im Hauptinterview des GSS die 13-Item-Version der Berufswerte (NORC-Y, Spalte 3) kategorialskaliert haben. Die in beiden Fällen beträchtliche Variation der Einzelwerte legt die Vermutung auf Ausreißer nahe und die Strategie, diese für den Vergleich zu eliminieren. Dies ist parallel für die ZUMA- und die NORC-Population in den Spalten 4 bzw. 5 geschehen. Hier wurden alle Befragten nicht berücksichtigt, die 1. eine negative Zwischenmodalitäten-Korrelation aufwiesen (in der NORC-Stichprobe fanden sich Korrelationen in der Größenordnung von $-.8!$), die 2. einen

ICMM-Exponenten aufwiesen, der signifikant vom erwarteten Wert 1 abwich (vgl. Tabelle 15, Spalten 7 und 8) und 3. deren Skalenbreiten (Range; vgl. den unteren Teil von Tabelle 11) außerhalb des Bereichs 2-200 fielen.

Bei Zugrundelegung dieser Kriterien verblieben für die ZUMA-Population noch 78.8% der Befragten, für die NORC-Population noch 58.5%. Während die mittleren Skalenwerte sich bei diesen Eliminierungen zwar angleichen, sind diese unterschiedlichen Prozentsätze doch ein Hinweis darauf, daß die Skalierungsgüte der Magnitude-Messungen in Amerika hinter der deutschen zurückbleibt. Dieses Resultat muß aber vor dem Hintergrund der unterschiedlichen Vertrautheit mit dem Verfahren gesehen werden. Die Interviewer der ZUMA-Untersuchung waren mit der Methode alle wohl vertraut; zum Teil hatten sie an der jahrelangen Entwicklungsarbeit des Verfahrens mitgewirkt. Die GSS-Interviewer dagegen arbeiteten zum ersten Mal mit der Magnitude-Methode. Sie wurden darüber hinaus lediglich schriftlich geschult, waren lokal nicht an das federführende Institut angebunden und hatten nicht die Möglichkeit, im Bedarfsfall mit den Verantwortlichen dieses Instituts Kontakt zur Klärung ihres Verständnisses aufzunehmen. Schließlich stellt die Umsetzung der bei ZUMA entwickelten Prozedur den ersten, nicht an Alternativversionen ausgetesteten Versuch dar, das Verfahren in den USA einzuführen. Bislang gibt es keine vergleichbaren Probeuntersuchungen, inwieweit das in der Bundesrepublik erfolgreiche Vorgehen auch für amerikanische Interviewer angemessen ist. Unter diesem Blickwinkel dürfen die amerikanischen Resultate der Erhebung in bezug auf die technische Güte der Magnitude-Messungen nicht zu negativ gesehen werden. Im Gegenteil, der noch relativ hohe Prozentsatz angemessener Skalierungen muß unter den erschwerten Bedingungen sogar als Kooperationserfolg gesehen werden.

Mit Ausnahme der Fälle mit negativer ICMM-Korrelation wurden alle Befragten in die späteren statistischen Analysen mit einbezogen, da sich die Magnitude-Skalierung auch unter faktischen Gesichtspunkten zu bewähren hat.

Weitere Verfahren zur Beurteilung der Skalierungsgüte von Magnitudemessungen:

Wie bereits erwähnt (vgl. Abschnitt 1.2 und oben) gibt es zwei weitere Möglichkeiten, die Skalenqualität von Magnitude-Messungen zu überprüfen:

- Axiomatische Analysen;
- Bestimmung der individuellen Urteolfunktionen.

A. Axiomatische Analysen

Während die vor allem an den drei oben erwähnten Kriterien gemessene numerische Skalierungsqualität der Magnitude-Skalen sich in bezug auf beide Teilstudien bestimmen ließ, weil in beiden direkte Magnitude-Messungen vorgenommen wurden, ist die Untersuchung der axiomatischen Skaleneigenschaften allein auf die deutsche Erhebung angewiesen: Zur meßtheoretischen Rekonstruktion der Skalen - der Magnitude- und auch der kategorischen - müssen zusätzlich zu den direkten Skalenwerten Paar-Stimulusschätzungen vorliegen. Die in Abschnitt 1.3 geschilderten Abstriche am gemeinsamen Forschungsprogramm ermöglichten solche Erhebungen nur beim ZUMA-Teilprojekt.

In vergleichbaren Untersuchungen (ORTH 1982; WEGENER 1983; ORTH & WEGENER 1983) hat sich erwiesen, daß Magnitude-Messungen in der Regel zu logarithmischen Intervallskalen und Kategorialskalierungen in der Regel zu Intervallskalen führen. Allerdings ergeben sich große interindividuelle Differenzen der realisierten Skalengüte. Man kann nicht davon ausgehen, daß die Anwendung eines der beiden Verfahren automatisch ein bestimmtes Skalenniveau sichert. Vielmehr verfügen unterschiedliche Befragte über unterschiedliche Skalierungskompetenzen, anhand deren sich bestimmte Befragtentypologien abgrenzen lassen.

Notwendige Bedingung für das Vorliegen einer logarithmischen Intervallskala im Fall der Magnitude-Messungen und einer Intervallskala im Fall der kategorischen Messungen ist, daß den individuellen Paar-Verhältnisschätzungen (im Magnitude-Fall) und den individuellen Paar-Differenzschätzungen (im kategorischen Fall), die beide im deutschen Teil der Studie erhoben wurden, jeweils eine Differenzen-Meßstruktur zugrundeliegt (vgl. z.B. ORTH 1974). Die algebraische Theorienbildung bezieht sich allerdings nur auf eindimensionale Stimuli und de facto wurde das Vorliegen derartiger Meßstrukturen bislang nur an eindimensionalem Reizmaterial überprüft. Dagegen erwies sich das Material der vorliegenden Untersuchung - vor allem im kategorischen Fall - als deutlich mehrdimensional. Dies ergaben nicht nur die auf aggregiertem Niveau durchgeführten Faktorenanalysen (vgl. Abschnitt 4.4), sondern auch die auf Individualniveau durchgeführten

nicht-metrischen mehrdimensionalen Skalierungen der Paar-Verhältnis- und Paar-Differenzschätzungen. Die Ergebnisse lassen sich an den individuellen Streßkoeffizienten dieser Analysen ablesen. Wir finden im Mittel über alle Befragten der deutschen Untersuchung für die Analyse der Paar-Differenzschätzungen einen Wert von .277 (SD=.109) und für die Analysen der Paar-Verhältnisschätzungen einen Mittelwert von .273 (SD=.070). Dies sind die Ergebnisse für die eindimensionale Rekonstruktion. Im Sinne von KRUSKAL (1964) können diese mittleren Streßwerte gerade noch als zufriedenstellend bezeichnet werden. Ihre Höhe ist aber ein deutlicher Hinweis dafür, daß wir es nicht mit einem eindimensionalen Stimulusraum zu tun haben. In zukünftigen Analysen ist aus diesem Grund darauf zu achten, daß die axiomatischen Meßstrukturen jeweils nur in bezug auf jene Stimuli untersucht werden, die getrennte Dimensionen etablieren bzw. beschreiben.

Die offensichtlich vorliegende Mehrdimensionalität, die im Fall der kategorischen und der Magnitude-Messungen durch jeweils unterschiedliche Dimensionen konstituiert wird (vgl. Abschnitt 5.3.1), ließ es nicht angebracht erscheinen, die Überprüfung der Axiomatik auf alle 13 Items gemeinsam auszudehnen, sondern nur auf jene, die eine einheitliche Skala unter den beiden untersuchten Methodenbedingungen bilden. Wie die weiteren Analysen noch zeigen werden, gilt dies lediglich für die statusbezogenen Unterskala, die sich aus den vier Items 'Sichere Berufsstellung', 'Hohes Einkommen', 'Aufstiegsmöglichkeit' und 'Anerkennung und Achtung' zusammensetzt. Tabelle 16 zeigt die mittleren Verstöße gegen die eine Differenzen-Meßstruktur kennzeichnende Quadrupelbedingung für die Differenzen- und Verhältnisschätzungen. Wie man sieht, sind die entsprechenden Werte ausgesprochen niedrig. Dies ist ein starker Indikator dafür, daß die kategorischen und Magnitudemessungen die jeweils erwarteten Skalentypen repräsentieren, selbst wenn man hinnehmen muß, daß die in der Regel über KENDALLS tau überprüfte Gültigkeit der Kompatibilitätsbedingung aufgrund der geringen Itemzahl nicht verläßlich ermittelt werden konnte.

B. Individuelle Urteilsfunktionen

Die Methodik der Bestimmung individueller Urteilsfunktionen wurde bereits in Abschnitt 1.2; S.6) genauer dargestellt. Die technische Voraussetzung für

die Anwendung dieser Methodik ist allerdings, daß die Anzahl der Stimuli genügend groß ist, um statistische Anpassungstests zu ermöglichen. Auch in diesem Zusammenhang wird wieder das Dimensionalitätsproblem relevant.

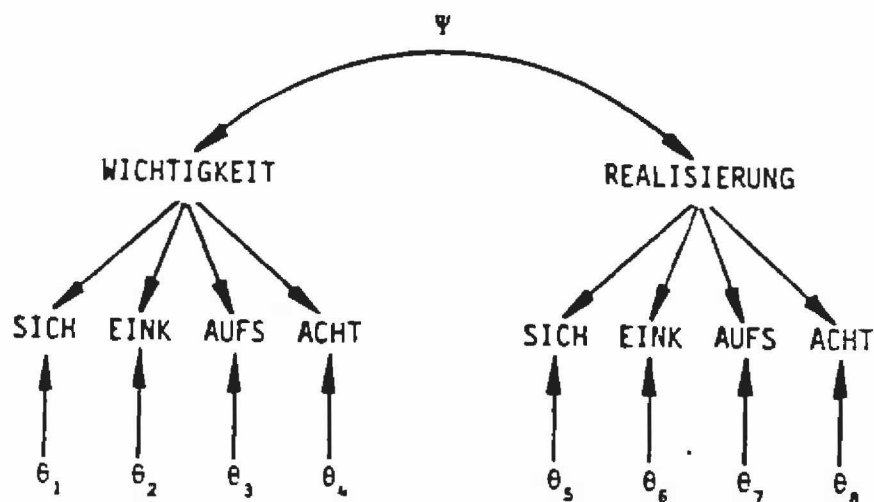
Tab. 16: Überprüfung der Quadrupelbedingung⁺

<u>Differenzenschätzungen</u>	<u>Verhältnisschätzungen</u>
.214	.223
(SD = .103)	(SD = .121)

⁺ $ab \geq cd \Rightarrow ac \geq bd$

Der Sachverhalt, daß lediglich die vier oben genannten Items einen vergleichbaren Faktor konstituieren, ließ es nicht sinnvoll erscheinen, die Urteilsfunktionen in bezug auf alle Befragte zu bestimmen. Als Alternative wurden die in Abschnitt 1.2 beschriebenen konfirmatorischen Faktorenanalysen mit den beiden Konstrukten 'Wichtigkeit allgemein' (kurz: WICHTIGKEIT) und 'Ausmaß der Realisierung im eigenen Beruf' (kurz: REALISIERUNG) in bezug auf die vier extrinsischen Items auf Aggregationsniveau gerechnet (vgl. Abbildung 11). Tabelle 17 gibt die Ergebnisse der Anpassungen wieder und macht deutlich, daß das zweifaktorielle Modell unter beiden Skalierungsbedingungen sehr gut rekonstruiert werden kann (GFI-Werte von .929 bzw. .920). Da die Magnitude-Skalenwerte der individuellen Befragten in die Analyse in logarithmierter Form eingingen, besagt das Ergebnis, daß wir es in diesen Fällen mit potenzförmigen Urteilsfunktionen zu tun haben.

Abb. 11: Schema des aggregierten Urteilsmodells für die vier extrinsischen Berufswerte



Dies gilt in gleicher Weise für die Items unter ihrem 'Wichtigkeitsaspekt' und (.W) und ihrem 'Realisationsaspekt' (.R). Auch aus diesem Ergebnis ergibt sich - zumindest für den deutschen Teil der Studie - ein weiterer Hinweis auf die Übereinstimmung der Magnitude- und kategorischen Messungen mit den theoretischen Grundlagen.

Allerdings fällt auf, daß sich die Reliabilitäten der Magnitude- und der kategorischen Messungen sich voneinander unterscheiden: Die Magnitude-Werte weisen höhere Reliabilitäten auf als die kategorischen Werte, wobei diese Reliabilitäten für die Beurteilung der 'Wichtigkeiten' jeweils höher sind als für die der 'Realisierungen'. Dies deutet darauf hin, daß die Urteilsfunktionen im Fall der Magnitude-Messungen eine größere Homogenität aufweisen und zwar in stärkerem Maß für die Wichtigkeitseinstufungen als für die Beurteilung der Berufswirklichkeit. Ein weiterer Unterschied zwischen kategorialen und Magnitude-Messungen betrifft den Zusammenhang der Konstrukte: Im Fall der

Tab. 17: Modellparameter der aggregierten Urteilsmodelle für die vier extrinsischen Items

	MAGNITUDE (LOG)				KATEGORISCH			
	LAMBDA.W	LAMBDA.R	THETA ²	RELIA	LAMBDA.W	LAMBDA.R	THETA ²	RELIA
1 SIC.H.W	.814	--	.337	.663	.637	--	.594	.406
2 EINK.W	.793	--	.372	.620	.752	--	.435	.565
3 AUFS.W	.934	--	.127	.873	.862	--	.258	.742
4 ACHT.W	.810	--	.345	.655	.652	--	.575	.425
5 SIC.H.R	--	.709	.497	.503	--	.423	.821	.179
6 EINK.R	--	.602	.638	.362	--	.611	.627	.373
7 AUFS.R	--	.816	.335	.665	--	.712	.493	.507
8 ACHT.R	--	.812	.340	.660	--	.619	.617	.383
	PSI = .760			.626	PSI = .483			.448
	CHI ² = 33.50; df = 19 (P = .021)				CHI ² = 38.88; df = 19 (P = .005)			
	CHI ² /df = 1.77				CHI ² /df = 2.05			
	GFI = .929				GFI = .920			
	RMSR = .048				RMSR = .070			
	RELIA.W = .705				RELIA.W = .535			
	RELIA.R = .540				RELIA.R = .361			

Magnitude-Messungen ergibt sich eine erheblich größere Konstruktkorrelation (.76) als bei den kategorischen Messungen (.48). Da wir es im letzteren Fall mit niedrigeren Reliabilitäten zu tun haben und außerdem mit einem kongenerischen, d.h. eindimensionalen Meßmodell, bleibt nur der Schluß, daß die kategorialen Messungen etwas vortäuschen, was man als 'Unzufriedenheitspotential' bezeichnen könnte: Die Berufswerte, die man für wichtig hält, findet man im eigenen Beruf nur unzureichend verwirklicht. Möglicherweise kann diese Konsequenz als Artefakt der Uneinheitlichkeit der Urteilsfunktionen von Kategorialeurteilen interpretiert werden.

5.2.2 Die kategorialen Messungen

Einen Anhaltspunkt für die Güte der kategorialen Messungen ergeben die Analysen der Interskalenrelationen (vgl. S.25). Letztere ermöglichen nämlich eine Validierung der Kategorialskalierung an der Magnitudeskalierung über die Verifikation der Form der funktionalen Beziehung. Die Parameter dieser Beziehung sind Tabelle 18 zu entnehmen. In dieser Tabelle ist mit RKAPPA die Korrelation der optimalen (linearisierten) Funktionsanpassung bezeichnet; RINF ist die Korrelation, die sich bei Anpassung eines alternativen logarithmischen Interskalenmodells ergibt; EXPO ist der Wert des Interskalenexponenten. Im oberen Teil der Tabelle sind diese Werte für den aggregierten Fall, im unteren für den individuellen (arithmetisch gemittelten) Fall aufgeführt (Standardabweichungen in Klammern).

Während die deutschen Resultate in Obereinstimmung stehen mit Ergebnissen, die in einer Vielzahl anderer Untersuchungen und anderen Stimulusmaterialien gefunden wurden (vgl. WEGENER 1983), und zwar sowohl für die aggregierte Rechnung (oberer Tabellenteil) als auch für Individualanalysen (unterer Tabellenteil), ist die Bestimmung der Interskalenrelationen bei der amerikanischen Population unbefriedigend. Die mittleren Individualergebnisse weichen signifikant von der deutschen ab, und zwar sogar dann, wenn jene Fälle eliminiert werden, deren Magnitudeskalen offensichtlich degenerativ sind (z.B. negative ICMM-Korrelationen aufweisen).

Insgesamt ergibt sich aus der Analyse der Interskalenrelationen, daß die amerikanischen Messungen offensichtlich nicht in dem Maße mit den theoretischen Modellannahmen einer logarithmischen Intervallskala (im Magnitudefall) und einer Intervallskala (im kategorialen Fall) verträglich sind wie die deutschen Messungen. Die möglichen Gründe hierfür wurden bereits im letzten Abschnitt (S. 51) diskutiert.

Auch die Ergebnisse der axiomatischen Analysen und der Bestimmung der Urteilsfunktionen im deutschen Teil der Untersuchung wurden bereits im letzten Abschnitt in einer vergleichenden Betrachtung mit den Magnitudeergebnissen dargestellt. Sie bestätigen die gute Kompatibilität der kategorialen Messungen mit theoretischen Modellannahmen zumindest im Fall der vier extrinsischen Items der Berufswerte-Ska-

Tab. 18: Parameter der Interskalenrelationen

INTERSKALEN INDICES	ZUMA	NORC-Y
<u>AGGR</u>		
RKAPPA	.873	.799
RINFIN	.870	.797
EXPO	.198	.330
N	103	56
<u>INDIV</u>		
RKAPPA	.739 (.825 ⁺) (.221)	.592* (.635 ⁺) (.264)
RINFIN	.703 (.790 ⁺) (.251)	.525* (.529 ⁺) (.301)
EXPO	.548 (.843)	.667 (1.198)
N	88	48

⁺ Median

* $p \leq .001$ verglichen mit ZUMA

5.3 Konstruktidentifikation und Konstruktvergleiche

5.3.1 Faktoren und Faktorstrukturen im Vergleich

Um einen Überblick über die Konstrukte zu erhalten, die den mittels der beiden Schätzverfahren (Magnitude und kategorial) in beiden Nationen erhobenen Wichtigkeitseinstufungen der Berufswerte zugrundeliegen, wurden zunächst entsprechende exploratorische Faktorenanalysen durchgeführt, deren Ergebnisse in Tabelle 19 dargestellt sind.

Tab. 19: Ergebnisse der exploratorischen Analysen

a. ZUMA

	<u>KATEGORISCH</u>				<u>LOG MAGNITUDE</u>			
	F1	F2	F3		F1	F2	F3	
SICHER	.73	-.03	.27		.14	.55	.67	
EINKOM	.86	.01	-.18		.37	.22	.83	
AUFSTG	.85	.21	.11		.27	.35	.80	
GEACHT	.69	.05	.28		.16	.30	.78	
FRZEIT	.01	.44	-.41		.90	.00	.20	
INTER	-.07	.74	.04		.84	.32	.22	
SELBST	.04	.82	.06		.76	.22	.36	
VERANT	.22	.71	.22		.38	.67	.37	
KONTKT	.14	.54	.38		.55	.60	.12	
HELFEN	.20	.01	.84		.24	.85	.12	
NUETZL	.34	-.06	.74		.13	.79	.33	
SINNVL	-.02	.35	.71		.79	.36	.17	
BEDING	.02	.22	.42		.68	.42	.19	
Varianz	29.2%	16.6%	12.9%	58.7%	55.9%	11.4%	9.1%	76.4%

b. NORC

	<u>KATEGORISCH</u>				<u>LOG MAGNITUDE</u>			
	F1	F2	F3		F1	F2	F3	
SICHER	.23	.37	.57		.15	.91	.20	
EINKOM	.38	.65	.25		.68	.13	.53	
AUFSTG	.38	.65	.12		.82	.45	.03	
GEACHT	.32	.65	.21		.75	.43	.19	
FRZEIT	-.27	.81	.19		.22	.36	.87	
INTER	.46	.32	.55		.83	.35	.31	
SELBST	.74	.20	.14		.85	.14	.28	
VERANT	.84	.08	.31		.89	.33	.16	
KONTKT	.03	.07	.89		.58	.66	.27	
HELFEN	.31	.38	.76		.62	.67	.24	
NUETZL	.37	.29	.75		.71	.56	.25	
SINNVL	.52	.25	.49		.81	.40	.25	
BEDING	.23	.71	.32		.51	.70	.31	
Varianz	48.6%	10.3%	8.0%	66.9%	73.7%	7.6%	5.2%	86.5%

Die Analysen stellen Hauptkomponentenlösungen ohne iterative Kommunalitätsschätzungen dar. Wiedergegeben sind immer die ersten drei Faktoren und der Anteil der durch sie aufgeklärten Varianz. Die Analysen beziehen sich im Fall der Magnitude-Messungen auf die logarithmierten Werte. Die Logarithmierung der Magnitudewerte erscheint im Kontext des Vergleichs mit den kategorialen Einstufungen darum sinnvoll, weil zwischen beiden Skalierungsformen potenzförmige Interskalenrelationen angenommen werden müssen. Bei der Verwendung der originalen Magnitudewerte hätten die von uns angestellten linearen statistischen Analysen automatisch zu einer Erhöhung des Meßfehlers geführt. Im übrigen haben Faktorenanalysen mit den ursprünglichen Magnitudewerten, die hier nicht wiedergegeben sind, gezeigt, daß sich die Faktorenstrukturen der logarithmierten Werte von denen der unlogarithmierten kaum unterscheiden.

Die Diskussion der faktorenanalytischen Ergebnisse im Falle der kategorialen Einstufungen der ZUMA-Teilstudie wurde bereits an anderer Stelle geführt (vgl. Abschnitt 4.3.1). Die drei Faktoren wurden dort als Extrinsische Orientierung, Intrinsische Orientierung und Soziale Orientierung bezeichnet. Rein aufgrund der visuellen Inspektion fällt die hohe Ähnlichkeit zwischen dem Faktor F3 der Faktorenanalyse der Magnitudewerte mit dem Faktor Extrinsische Orientierung (F1) der Faktorenanalyse der kategorialen Werte auf. Das gleiche gilt für die Ähnlichkeit von F1 (Magnitude) und Intrinsische Orientierung (F2) sowie zwischen F3 (Magnitude) und Soziale Orientierung (F3) auf, sodaß man insgesamt versucht wäre schon auf der Basis der visuellen Inspektion den Schluß zu ziehen, daß beide Methoden zu sehr ähnlichen Faktorenstrukturen geführt haben.

Der Versuch, auf der Basis der visuellen Inspektion allein auf Ähnlichkeiten der Faktorenstrukturen beider Methoden im Fall der NORC-Teilstudie zu schließen gestaltet sich erheblich schwieriger. Ähnlich schwierig sieht es auch mit dem Versuch aus, die Faktorenstrukturen für jede Methode getrennt über beide Nationen hinweg zu vergleichen.

Ogleich schon bei einer ersten visuellen Inspektion der Faktormatrizen gewisse Ähnlichkeiten und Unterschiede ins Auge fallen, muß der visuellen Beurteilungsmethode doch mißtraut werden. Es kann dabei allzu leicht vorkommen, daß Ladungsunterschiede über- oder unterbewertet werden und die Vergleichsergebnisse daher ein verzerrtes Bild liefern. Besser erscheint es unter diesem Gesichtspunkt,

sich formaler Methoden des Vergleichs zu bedienen.

Zum Vergleich von Faktormatrizen stehen verschiedene Verfahren zur Auswahl (vgl. z.B. RUMMEL 1970). Die meisten dieser Verfahren folgen dem Vektoransatz. Dabei werden zwei Faktormatrizen faktorenweise miteinander verglichen, wobei das Ausmaß an Ähnlichkeit schließlich in gewissen Koeffizienten festgehalten wird (vgl. z.B. CATTELL 1957:821-828; HARMAN 1967:269-270). Die in dieser Studie angewandte Methode von AHMAVAARA (1954a, 1954b; vgl. auch RUMMEL 1970:463) besteht in einem Vergleich, nicht der Ladungsvektoren, sondern der Ladungsmatrizen. Wie RUMMEL (1965) gezeigt hat, kann diese Methode sehr sinnvoll in internationalen Vergleichsstudien angewandt werden.

Das von AHMAVAARA als Transformationsanalyse bezeichnete Verfahren besteht darin, daß zunächst eine Transformationsmatrix bestimmt wird, die die Faktoren der ersten Studie nach dem Kriterium der kleinsten Quadrate auf Faktoren der zweiten Studie rotiert, d.h. die die Kongruenz zwischen beiden Faktorenmengen maximiert. Die Elemente dieser Transformationsmatrix messen in normalisierter Form die Enge des Zusammenhangs zwischen den Faktoren beider Studien im gleichen Faktorraum. Über den Vergleich der transformierten Faktorenladungen der ersten Studie mit den Faktorenladungen der zweiten Studie lassen sich als globale Ähnlichkeitsmaße Korrelationskoeffizienten berechnen: PEARSONS r und der Intraklassen-Korrelationskoeffizient (vgl. KENDALL & STUART 1979). Rechentechnisch steht für dieses Verfahren das Programm FCOMP des Programmpakets OSIRIS III zur Verfügung.

Tabelle 20 enthält neben Ähnlichkeitsmaßen, die die in Tabelle 19 dargestellten Faktormatrizen einbeziehen, auch Maße für die Ähnlichkeiten einiger weiterer Matrizenpaare, die hier nicht gesondert abgebildet sind. Die in Tabelle 20 auftretenden Abkürzungen dürften ohne weitere Erläuterungen zu verstehen sein. NORC-Y/MAG bezieht sich auf die Faktormatrix der Magnitudewerte für jene Befragten der NORC-Stichprobe, die die Version Y der Frage nach der Wichtigkeit von Berufswerten beantworteten.

Tab. 20: Korrelationen zwischen den Faktorstrukturen

	<u>PEARSON</u>	<u>INTRAKLASSEN</u>
ALLBUS 1982 - GSS 1982	.793	.777
NORC/KAT - GSS 1982	.652	.615
ZUMA/KAT - ALLBUS 1982	.820	.820
ZUMA/MAG - ZUMA/KAT	.850	.842
NORC-Y/MAG - NORC/KAT	.909	.907
ZUMA/KAT - NORC/KAT	.853	.842
ZUMA/MAG - NORC/MAG	.704	.680

Eine Inspektion von Tabelle 20 ergibt relativ hohe Korrelationen zwischen den Faktormatrizen der beiden Methoden bei Konstanzhaltung der Nation. Bei den kategorialen Einschätzungen ergeben sich auch hohe Korrelationen zwischen den Nationen bei Konstanzhaltung der Methode. Im Vergleich hierzu fallen die Korrelationen zwischen den Magnitude-Matrizen der beiden Nationen deutlich ab. Dies erscheint aufgrund der bereits in Abschnitt 5.2.1 diskutierten minderen Qualität der NORC-Magnitude-Skalierungen durchaus plausibel.

Bei der Beurteilung der Repräsentativität der Ergebnisse verdienen die hohen Korrelationen zwischen den Matrizen der ZUMA-Kategorialskalierung und des ALLBUS 1982 besondere Beachtung. Sie liegen erheblich höher als die Korrelationen zwischen den Matrizen der NORC-Kategorialskalierung und dem GSS 1982. Letztere liegen sogar unter den Korrelationen zwischen den Matrizen von ALLBUS 1982 und GSS 1982.

Die betrachteten Korrelationen stellen globale formale Ähnlichkeitsmaße dar. Will man wissen, inwieweit einzelne Faktoren der verglichenen Matrizen Ähnliches messen, erscheint es sinnvoll, die normalisierten Transformationsmatrizen der einzelnen Vergleiche zu betrachten. Die in diesen Matrizen enthaltenen Werte sind als Maßzahlen für die Enge des linearen Zusammenhangs zwischen den Faktoren der Studie 1 (Zeilen) und den Faktoren der Studie 2 (Spalten) zu interpretieren. Sie können zwischen +1.00 und -1.00 variieren.

Tab. 21: Ähnlichkeiten zwischen den Faktoren

		<u>ZUMA/KAT</u>		
		<u>F1</u>	<u>F2</u>	<u>F3</u>
<u>ZUMA/MAG</u>	<u>F1</u>	0.40	0.90	0.15
	<u>F2</u>	0.08	0.03	1.00
	<u>F3</u>	0.98	0.11	0.19

		<u>NORC/KAT</u>		
		<u>F1</u>	<u>F2</u>	<u>F3</u>
<u>NORC-Y/MAG</u>	<u>F1</u>	0.96	0.17	0.22
	<u>F2</u>	0.16	0.99	0.04
	<u>F3</u>	0.20	0.39	0.90

		<u>NORC/MAG</u>		
		<u>F1</u>	<u>F2</u>	<u>F3</u>
<u>ZUMA/MAG</u>	<u>F1</u>	0.69	0.24	0.68
	<u>F2</u>	0.55	0.83	0.10
	<u>F3</u>	0.88	0.34	0.32

		<u>NORC/KAT</u>		
		<u>F1</u>	<u>F2</u>	<u>F3</u>
<u>ZUMA/KAT</u>	<u>F1</u>	0.41	0.85	0.32
	<u>F2</u>	0.72	0.47	0.51
	<u>F3</u>	0.44	0.10	0.89

Für die Vergleiche über die Nationen hinweg unter Konstanthaltung der Nation und über die Nationen hinweg bei Konstanthaltung der Methode sind die Ergebnisse der Faktorenvergleiche in Tabelle 21 wiedergegeben. Besonders interessant sind jene Faktoren der Zeilenmatrix, für die es genau einen Faktor der Spaltenmatrix gibt, mit dem dieser Faktor sehr hoch korreliert ist. Ein Beispiel ist etwa der Faktor F3 von ZUMA/MAG, der zu .98 mit F1 von ZUMA/KAT korreliert, aber nur zu .11 und .19 mit den übrigen Faktoren von ZUMA/KAT. Ein so strukturierter Zusammenhang ist wichtig, weil der Hinweise für eine ähnliche Interpretation der Faktoren abgeben kann. Auf diese Weise ergeben sich z.B. Hinweise dafür, daß die Methoden der Magnitude-Messung und der kategorialen Messung sowohl im Fall der ZUMA-Teilstudie als auch im Fall der NORC-Teilstudie zu ähnlich interpretierbaren Faktoren führen, während derartige Hinweise für die Vergleiche über die Nationen hinweg nicht so eindeutig ausfallen.

Bei der Interpretation korrelativer Zusammenhänge zwischen Faktoren muß allerdings bedacht werden, daß Korrelationen zwischen Faktorenladungen auf linearen statistischen Analysen beruhen und damit die zulässigen Transformationen, durch die die Ähnlichkeit der zu vergleichenden Konfigurationen maximiert werden soll, auf die Klasse linearer Transformationen beschränkt bleibt. Rotationen und Transformationen sind keineswegs die einzigen Transformationstypen, die auf die Punkte von Strukturen anwendbar sind. Unter Umständen lassen sich Strukturähnlichkeiten nur entdecken, wenn man zusätzliche Skalenstauchungen und -streckungen oder sogar perspektivische Punkteverlagerungen vornimmt. Derartige Transformationstypen fallen wegen der metrischen Annahmen für faktorenanalytisch gewonnene Strukturen aus. Sie kommen erst in bezug auf Konfigurationen in Frage, die aus schwächeren Voraussetzungen entstanden sind, insbesondere in bezug auf Konfigurationen, die bloße Distanzmodelle darstellen. Neben Rotationen und Reflexionen sind solche Strukturen linear beliebig transformierbar, ihr Ursprung ist variabel, sie können dimensionsspezifisch gewichtet und als ganzes umplaziert werden.

Eine weitere Beschränkung faktorenanalytischer Vergleiche liegt darin, daß der Prozentsatz der Varianzaufklärung unberücksichtigt bleibt. So zeigt z.B. ein Blick auf Tabelle 19, daß in allen Fällen, die Faktorenanalysen der Magnitude-Messungen mehr Varianz aufklären als die Faktorenanalysen der entsprechenden kategorischen Daten und daß vor allem der erste Faktor der Magnitude-Faktorenanalysen im Verhältnis zu den übrigen Faktoren erheblich mehr Varianz aufklärt. Die liegt die Vermutung nahe, daß die Magnitude-Skalierung möglicherweise

dimensionsreduzierend wirkt. Die Frage, die sich in diesem Zusammenhang stellt, ist die, wie ein Vergleich zwischen beiden Meßmethoden zu beurteilen ist, der jeweils unterschiedliche Faktoranzahlen zuläßt.

5.3.2 Ergebnisse der PINDIS-Analysen

Ein Verfahren, welches den Vergleich von Konfigurationen unter weniger restriktiven Transformationsbedingungen und unter Zulassung unterschiedlicher Anzahlen von Dimensionen gestattet ist ein unter dem Namen PINDIS bekanntes Verfahren der multidimensionalen Skalierung (vgl. LINGOES & BORG 1976; BORG & LINGOES 1978; BORG 1977). Bei PINDIS handelt es sich den umfassendsten Algorithmus für die Bearbeitung von Strukturvergleichen. Frühere Versuche, die in dieselbe Richtung gehen, sind das INDSCAL- und das IDIOSCAL-Modell (CARROLL & WISH 1974), die aber beide die Freizügigkeit der zulässigen Transformationen noch begrenzen. Erst PINDIS vereint alle denkbaren Transformationen und arbeitet sie in der Form eines hierarchischen Modells solcher Transformationen ab.

Die Prozedur nimmt ihren Ausgang von einer Menge individueller Konfigurationen X_i ($i=1, \dots, n$), die zur Verfügung stehen und z.B. das dimensionsgleiche Ergebnis von multidimensionalen Analysen in bezug auf identische Stimuli sein können. Nachdem die Einzelkonfigurationen normiert wurden, d.h. einen gemeinsamen Ursprung erhalten haben und dimensional so transformiert wurden, daß ihr "allgemeiner Skalenfaktor" (Quadratwurzel der Summe der Quadrate der Koordinatenwerte) 1 ist, wird die Durchschnittskonfiguration Z der einzelnen Konfigurationen gebildet. Z bestimmt sich aus den mittleren Koordinatenwerten der Einzelkonfigurationen, nachdem diese orthogonal in der Weise rotiert und reflektiert wurden, daß die Summe der quadrierten Distanzen zwischen korrespondierenden Punkten X_i ein Minimum bilden (ordinary least squares). Auf diese Weise stellt die Durchschnittskonfiguration Z jene Konfiguration dar, die allen X_i nach einfachen zulässigen Transformationen gemeinsam ist.

Für jede individuelle Konfiguration versucht PINDIS das Ausmaß der Übereinstimmung mit Z zu quantifizieren. Dies geschieht formal durch die Angabe des quadrierten Korrelationskoeffizienten zwischen den korrespondierenden Koordinatenwerten von Z und X_i in bezug auf alle Punkte; d.h. $r^2(Z, X_i)$ drückt die Kommunalität zwischen der Durchschnittskonfiguration und einer individuellen Konfiguration

X_j aus bzw. - in regressionsanalytischer Ausdrucksweise - den Anteil der Varianz, in X_j , der durch Z erklärt wird.

An dieser Stelle setzt nun der Prozeß der idiosynkratischen Transformationen an. Die Kommunalitäten werden nicht einfach in bezug auf Z als gemittelte Zentroidkonfigurationen berechnet, sondern sie werden in einem hierarchischen Transformationsprozeß, dem Z unterworfen wird, optimiert. Zur Bestimmung der maximalen "konfiguralen Ähnlichkeit" zwischen Z und X_j wird Z sequentiell sechs Formen von Transformationen ausgesetzt, die unter der Bedingung zulässig sind, daß lediglich die relativen Verhältnisse der Interpunkt-Distanzen bewahrt werden. Jede dieser Transformationen wird so vorgenommen, daß die Kommunalität zwischen Z und einem X_j in bezug auf diese Transformation maximal wird. Auf diese Weise drückt r^2 die maximale Ähnlichkeit zwischen Z und X_j aus, die entsteht, wenn alle Freiheitsgrade der Skalentransformationen ausgeschöpft sind.

Es handelt sich um folgende Transformationen, die sukzessiv an Z durchgespielt werden:

1. Z (mit X_j auf die Euklidische Norm 1 gebracht) wird lediglich rotiert/reflektiert, um mit den Punkten X_j optimal zur Deckung zu kommen: $r^2(Z, X_j)$.
2. Z wird dimensionsweise gewichtet: $r^2(ZW_j, X_j)$.
3. Z wird dimensionsweise gewichtet, aber vorher (wie bei 1.) auf eine optimale Position rotiert/reflektiert: $r^2(Z_jW_j, X_j)$.
4. Z wird vektorweise gewichtet, und zwar in bezug auf jeden Ursprung-Punkt-Vektor der Konfiguration, und X_j wird vorher optimal auf Z rotiert/reflektiert: $r^2(V_jZ, X_j)$.
5. Z wird vektorweise gewichtet, aber als ganzes translatiert (perspektivische Ursprungsverschiebung), und X_j wird vorher optimal auf Z rotiert/reflektiert: $r^2(V_jZ_j, X_j)$.
6. Dimensionale und vektorweise Gewichtung werden simultan an Z vorgenommen, X_j wird zuvor auf Z rotiert/reflektiert: $r^2(V_jZW_j, X_j)$.

Unter der Voraussetzung dieser Klassen von Transformationen in bezug auf Z (es mag in Einzelfällen Gründe geben, auf eine oder mehrere dieser Transformationsarten zu verzichten) drücken die Kommunalitäten $r^2(Z, X_j)$, die kleiner als 1

sind, das Ausmaß konfiguraler Unähnlichkeit aus, die gewissermaßen übrigbleibt zwischen der gemeinsamen Durchschnittskonfiguration und einer individuellen Konfiguration. Der Grad dieser Nicht-Übereinstimmung läßt sich in Gewichtsform quantifizieren: Die in bezug auf ein X_j transformierte Konfiguration Z wird dimensionsweise gewichtet, um X_j zu erhalten. Die Gewichte w_{mj} drücken aus, in welcher Weise die m Dimensionen von Z gestreckt oder gestaucht werden müssen, um mit X_j identisch zu sein. Während also $r^2(Z, X_j)$ eine Kennzahl für die generelle Ähnlichkeit bzw. Unähnlichkeit einer individuellen Konfiguration mit Z ist, drücken die Gewichte w_{mj} aus, in bezug auf welche spezifischen Dimensionen die Ähnlichkeit bzw. Unähnlichkeit besteht.

Zum Verständnis der nachfolgenden Anwendung des Verfahrens auf unsere Fragestellung richten wir unser Augenmerk also auf 1. die Kommunalitäten, 2. die gemeinsame Konfiguration (Objektraum) und 3. auf den Raum der Gewichte (Subjektraum).

Im Rahmen einer PINDIS-Analyse haben wir die in Tabelle 22 angegebenen Gruppen bzw. Pseudo-Subjekte hinsichtlich der Wichtigkeitsbeurteilungen der Berufswerte miteinander verglichen.

Tab. 22: Pseudo-Subjekte der PINDIS-Analyse

	KATEGORISCH	MAGNITUDE
ZUMA	I	II
NORC	III	IV
ZUMA+NORC	V	VI

Als Grundlage dafür wurde für jede Gruppe eine Stimuluskonfiguration durch nicht-metrische multidimensionale Skalierung erstellt (smallest space analysis, SSA). In Tabelle 23 finden sich die Streß₁-Werte für die drei-, zwei- und eindimensionalen Lösungen. Die dreidimensionalen Konfigurationen wurden für

die PINDIS-Analyse zugrundegelegt. Der Streß₁-Wert entspricht dem von KRUSKAL (1964) eingeführten Streßkoeffizienten, der ein Maß für den Grad der Unangepaßtheit der multidimensionalen Lösungen darstellt (vgl. auch S.53 oben).

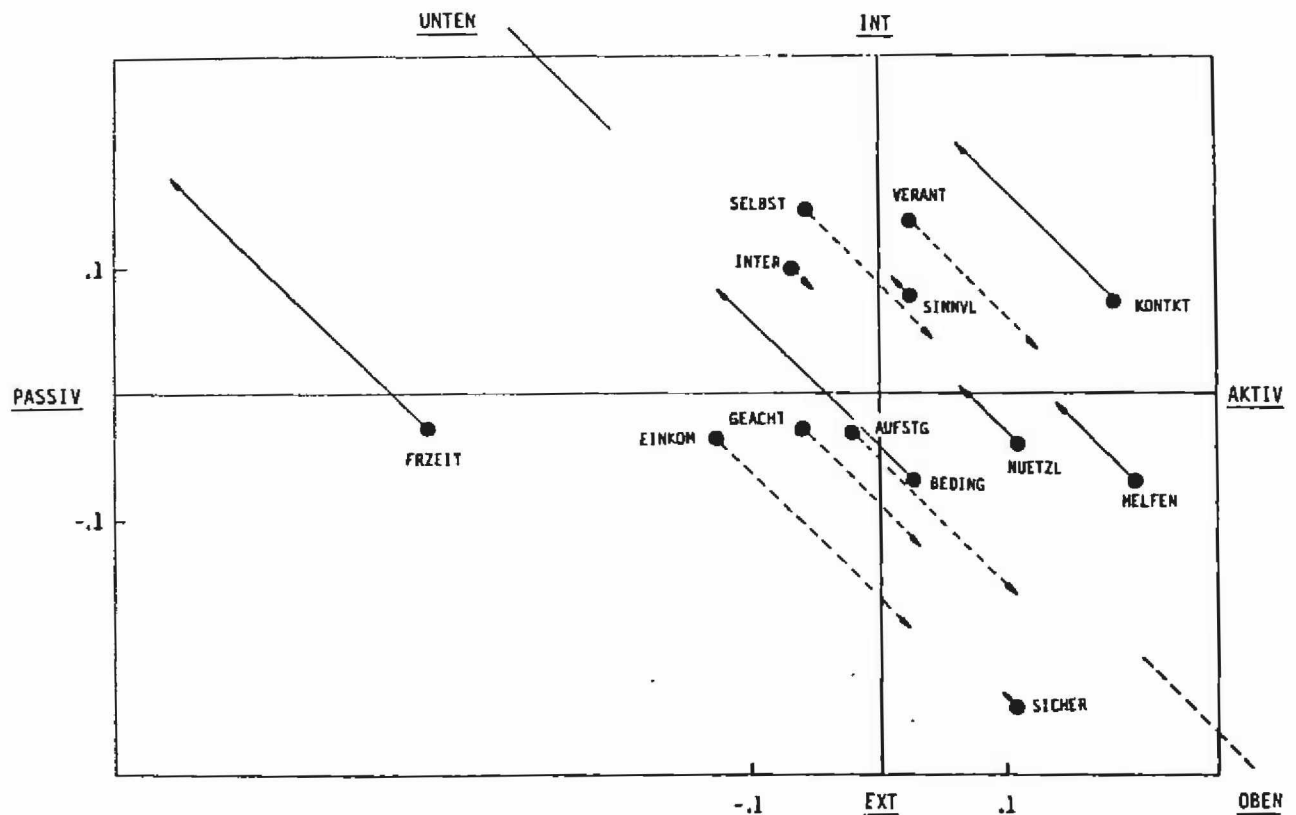
Tab. 23: Streß₁-Werte für die ein- bis dreidimensionalen Lösungen

		3 Dimensionen	2 Dimensionen	1 Dimension
ZUMA	Kategorisch	.078	.141	.445
	Magnitude	.070	.119	.330
NORC	Kategorisch	.103	.175	.356
	Magnitude	.010	.060	.113
ZUMA +NORC	Kategorisch	.065	.129	.205
	Magnitude	.010	.048	.125

Als Ergebnis der PINDIS-Analysen ergibt sich zunächst Tabelle 24 als Zusammenfassung der Kommunalitäten $r^2(Z, X_i)$, d.h. der generellen Ähnlichkeitsmaße der sechs Einzelkonfigurationen in bezug auf die allen gemeinsame Konfiguration Z. Die im Mittel größte Ähnlichkeit ergibt sich bei der vektoriellen Transformation mit perspektivischer Translation ($V_i Z_i, X_i$), die geringste Ähnlichkeit liegt bei einfacher Rotation bzw. Reflektion von Z vor (Z, X_i). Allerdings variieren diese Werte in bezug auf die sechs Gruppen. Entsprechend erwarten wir unterschiedliche Gewichtungsräume, die das Ausmaß der Ähnlichkeit/Unähnlichkeit spezifizieren.

Abbildung 12 ist die graphische Darstellung von Z, der gemeinsamen Objektkonfiguration. Die Abbildung ist der Versuch, die dreidimensionale Lösung, deren Koordinatenwerte in Tabelle 25 wiedergegeben sind, auf der zweidimensionalen Ebene darzustellen. Die dritte Dimension wurde zu diesem Zweck in Form diagonalen Vektoren veranschaulicht.

Abb. 12: Gemeinsame Objektkonfiguration



Wenn wir diese Diagonalen zunächst unbeachtet lassen, wird deutlich, daß die vertikale Dimension der Verteilung der Berufswerte auf dieser Dimension nach ein Kontinuum von intrinsischen und extrinsischen Berufswerten beschreibt. Die horizontale Dimension läßt sich demgegenüber als ein Kontinuum interpretieren, das Eigenschaften der beruflichen Tätigkeit selbst (z.B. HELFEN und KONTKT) von solchen trennt, die konsumptive Möglichkeiten eines Berufs bezeichnen (FRZEIT, EINKOM, GEACHT). Wir nennen entsprechend die vertikale Dimension die intrinsisch/extrinsische Dimension und die horizontale die Aktiv/Passiv-Dimension.

Die Plazierungen der Items auf der dritten Dimension werden durch die Richtungen der diagonalen Vektoren und ihre Länge markiert. Je länger die durchgezogenen

Tab. 24: Kommunalitäten

KONFIGURATION	TRANSFORMATION					
	Z, X_i	ZW_i, X_i	Z_iW_i, X_i	V_iZ, X_i	V_iZ_i, X_i	V_iZW_i, X_i
1	0.67	0.71	0.00	0.96	0.98	0.79
2	0.60	0.65	0.00	0.89	0.85	0.75
3	-0.58	0.58	0.00	0.85	0.85	0.71
4	0.73	0.76	0.00	0.91	0.97	0.81
5	0.80	0.82	0.00	0.95	0.98	0.88
6	0.78	0.80	0.00	0.97	0.99	0.84
Mittelwert	0.69	0.72	0.00	0.92	0.94	0.80

Tab. 25: PINDIS-Objektraum

SICHER	.111	-.244	.004
EINKOM	-.127	-.032	-.213
AUFSTG	-.019	-.031	-.182
GEACHT	-.062	-.026	-.130
FRZEIT	-.355	-.029	.280
INTER	-.067	.100	-.021
SELBST	-.055	.143	-.144
VERANT	.026	.136	-.134
KONTKT	.186	.073	.172
HELFEN	.203	-.067	.086
NUETZL	.109	-.036	.057
SINNVL	.027	.081	.014
BEDING	.022	-.069	.212

Pfeile eines Punktes sind, um so höher ist der Koordinatenwert dieses Punktes auf der dritten Dimension "links" vom Ursprung; je länger die gestrichelten Pfeile hingegen sind, um so höher ist der Wert eines Items "rechts" vom Ursprung auf dieser Dimension. Man kann sich auf diese Weise leicht deutlich machen, daß die "diagonale" Dimension der Objektkonfiguration eine allgemeine Oben-Unten-Dimension darstellt. Ihr einer Pol (gestrichelte Pfeile) wird markiert durch Berufseigenschaften wie Hohes Einkommen, Aufstiegsmöglichkeiten, Unabhängigkeit, Verantwortung und Prestige, d.h. von Merkmalen, die bei "höheren" Berufen anzutreffen sind. Der entgegengesetzte Pol (durchgezogenen Pfeile) hingegen vereinigt Berufsattribute von weniger "wichtigen" bzw. statusniedrigen Berufen wie z.B. Viel Freizeit, die Verwirklichung "sozialer Anliegen" (KONTKT, HELFEN, NUETZL) und die Beachtung von äußeren, nämlich gesunden und sicheren Arbeitsbedingungen (BEDING).

Wir können also die Objektkonfiguration, die allen Pseudo-Subjekten gemeinsam ist, als einen Raum interpretieren, der durch die drei Dimensionen AKTIV/PASSIV, INTRINSISCH/EXTRINSISCH und OBEN/UNTEN gebildet wird. Wir wissen, daß die individuellen Konfigurationen dieser Durchschnittskonfiguration nur mehr oder weniger ähnlich sind. Durch die Angabe von Gewichten für jede der drei Dimensionen wird spezifiziert, in welcher Weise die individuellen Konfigurationen von der gemeinsamen Konfiguration auf den drei Dimensionen abweichen.

Die errechnete (normierte) Gewichtsmatrix findet sich in Tabelle 26.

Tab. 26: PINDIS-Subjektraum

<u>DIMENSIONSGEWICHTE</u>				<u>(KOMMUNALITÄTEN)</u>
ZUMA KAT.	.549	.506	.391	(.711)
ZUMA MAG.	.545	.113	.586	(.654)
NORC KAT.	.490	.346	.468	(.579)
NORC MAG.	.384	.352	.699	(.760)
Z/N KAT.	.614	.472	.470	(.820)
Z/N MAG.	.459	.321	.698	(.801)
Mittelwert	.262	.140	.319	(.721)

Sie beschreibt den PINDIS-Subjektraum; jedes der sechs Pseudo-Subjekte erhält in diesem Raum auf den drei Gewichtsvektoren Koordinatenwerte. Für jeweils zwei der Gewichtsvektoren sind die Lokationen der Pseudo-Subjekte in den Abbildungen 13a-c veranschaulicht. Die als Punkte im Subjekt- oder Gewichteraum (deren Dimensionen auf 1 normiert sind) dargestellten Populations- und Methoden-
gruppen tragen die folgenden Bezeichnungen:

ZC für ZUMA KATEGORISCH
ZM für ZUMA MAGNITUDE
NC für NORC KATEGORISCH
NM für NORC MAGNITUDE
ZNC für ZUMA+NORC KATEGORISCH
ZNM für ZUMA+NORC MAGNITUDE

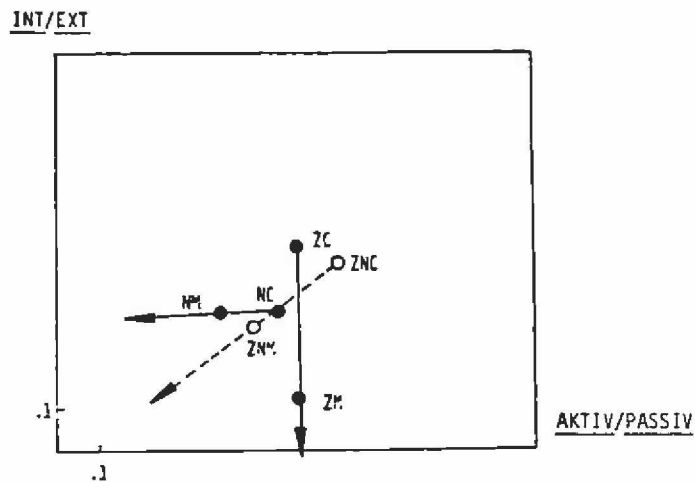
Die Interpretation des Subjektraums kann unter zwei Perspektiven erfolgen: Aufgrund von Interpunkt-Distanzen und dimensional. Die Interpunkt-Distanzen sind ein direktes Maß für die Unähnlichkeit zwischen den individuellen Konfigurationen (Der Cosinus zwischen zwei Ursprungs-Punkt-Vektoren ist numerisch die Korrelation zwischen den entsprechenden beiden Konfigurationen). Aus den Abbildungen 13a-c ist zu ersehen, welche Pseudo-Subjekte in bezug auf jeweils zwei Dimensionen nahe beieinanderliegen, d.h. ähnlich sind. Tabelle 26 läßt eine solche Betrachtung für alle drei Dimensionen simultan zu.

Inhaltlich aufschlußreicher ist die dimensionale Betrachtungsweise. Die Größe der Gewichte drückt die "Bedeutung" (salience) der entsprechenden Dimensionen für eine individuelle Konfiguration aus: Die gemeinsame Konfiguration ist nach Maßgabe dieser Gewichte dimensionsspezifisch zu strecken bzw. zu stauchen, um eine bestimmte individuelle Konfiguration rekonstruieren zu können. Wir können an dem Subjektraum also ablesen, wie bedeutungsvoll unterschiedliche Dimensionen für bestimmte Pseudo-Subjekte sind.

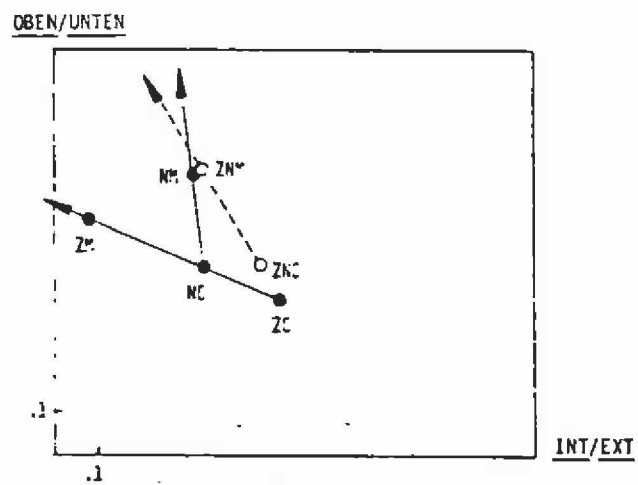
In diesem Sinne wird in den Abbildungen 13a-c für die Interpretation ein Leitfaden gegeben: Es sind dort Verbindungspfeile eingezeichnet, die jeweils zwei Pseudo-Subjekte miteinander zusammenfügen. Auf diese Weise verbunden sind jeweils die Punkte identischer Gruppen, aber unterschiedlicher Methoden. Mit anderen Worten, der Verbindungspfeil von ZC nach ZM z.B. zeigt die Richtung der Gewichtsveränderung an, die sich ergibt, wenn man an der ZUMA-Population die 13 Berufswerte

Abb. 13: Darstellung der Pseudo-Subjekte in jeweils zwei Dimensionen

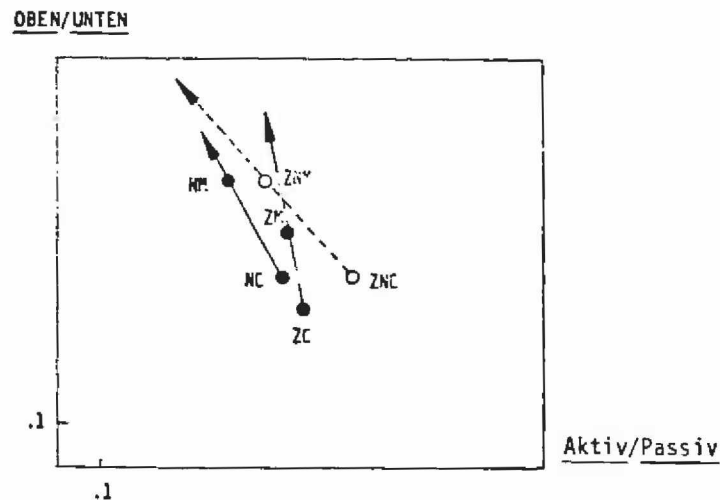
a. AKTIV/PASSIV vs. INT/EXT



b. INT/EXT vs. OBEN/UNTEN



c. AKTIV/PASSIV vs. OBEN/UNTEN



mit Magnitude-Technik statt mit Kategorialskalierung mißt. Die Punkte NC und NM bzw. ZNC und ZNM sind entsprechend verbunden. Ausgerüstet mit diesen Hilfslinien ist es leicht, an den Abbildungen die folgenden Ergebnisse abzulesen:

Abbildung 13a: Diese Abbildung stellt die Gewichtskoordinaten für die intrinsisch/extrinsische Dimension und die Aktiv/Passiv-Dimension dar. Wenn wir die Pfeile verfolgen, erkennen wir, daß für die ZUMA-Population ein dramatischer Abfall der Bedeutung der intrinsisch/extrinsischen Dimension zu beobachten ist, wenn die Skalierung als Magnitude- und nicht als kategorische Messung vorgenommen wird. Das Umgekehrte gilt für die NORC-Population: Der Übergang von der kategorischer zur Magnitude-Messung führt hier zu einer Reduktion der Bedeutung der Aktiv/Passiv-Dimension. Bemerkenswert ist in beiden Fällen, daß die jeweils anderen Dimension - die aktiv/passive bei ZUMA und die intrinsisch/extrinsische für NORC - von diesen Gewichtsverlagerungen unberührt bleibt. Außerdem ist zu beobachten, daß die gemeinsamen Populationen von ZUMA und NORC (offene Punkte) mit dem Methodenwechsel auf beiden Dimensionen Bedeutungsverluste erleiden.

Abbildung 13b: Die Gegenüberstellung der Oben/Unten-Dimension mit der instrinsisch/extrinsischen ergibt nach demselben Interpretationsverfahren, daß in diesem Fall die Oben/Unten-Dimension bei allen drei Populationen an Bedeutung gewinnt, wenn die kategorische Methode gegen das Magnitude-Verfahren ausgetauscht wird.

Für die amerikanische Stichprobe ist dieser Effekt allerdings stärker als für die deutsche. Die Wirkung auf die intrinsisch/extrinsische Dimension entspricht notwendigerweise der, die bereits an Abbildung 14a ablesbar war.

Abbildung 13c: Diese Abbildung drückt aus, was wir bereits kennen: den Bedeutungszuwachs in bezug auf die Oben/Unten-Dimension bei allen Gruppen und die differentiellen Auswirkungen auf die Aktiv/Passiv-dimension.

Die Schlußfolgerungen aus dieser Analyse des PINDIS-Subjektraums sind eindeutig. Die wesentlichste Feststellung lautet: Wenn man die Wichtigkeit von Berufswerten mit Magnitude-Methoden und nicht mit kategorischen Antwortvorgaben mißt, ergibt sich ein subjektiver Bedeutungszuwachs auf der generellen Oben/Unten- Dimension der Einschätzungen. Die anderen Dimensionen des Bewertungsraums werden demgegenüber in ihrer Bedeutung zurückgedrängt. Allerdings, so ist weiter festzustellen, sind von dieser Zurücknahme der Bedeutung in der deutschen und in der amerikanischen Population unterschiedliche Dimensionen betroffen. In der ZUMA-Population verliert die intrinsische/extrinsische Dimension an Gewicht. In der NORC-Population gilt dies für die Aktiv/Passiv-dimension. Beide Dimensionen sind für die kategorische Beurteilung der Berufswerte offenbar konstitutiv; sie treten - stichprobenspezifisch in den Hintergrund beim Übergang zur Magnitude-Methode der Erhebung.

Damit stehen wir vor dem Resultat, daß die Magnitude-Messungen offensichtlich dimensionsreduzierend wirken, zumindest bei den thematischen Einstellungen, die in dieser Untersuchung Forschungsgegenstand sind. Die Dimension, auf die reduziert wird, ist die generelle Oben/Unten-Dimension der Bewertung der Berufsattribute. Der Befund gilt für beide untersuchten Stichproben und konsequenterweise auch für die gepoolte Population. Mit diesem Ergebnis kommen wir unserer Leitfrage nach der Vergleichbarkeit der deutschen und amerikanischen Instrumente einen Schritt näher: Magnitude-Messungen rufen offensichtlich eine Faktorreduktion hervor, und zwar unabhängig von der Bevölkerung, die untersucht wird.

5.4 Die Identifikation von Indikatoren

Der Versuch, analog zu unserem Vorgehen im Fall des Vergleichs der numerischen und verbalen Versionen der Kategoriaskalierung (vgl. Kap. 3) über die

der Ergebnisse exploratorischer Faktorenanalysen (vgl. Tabelle 19; S.59) und die anschließende Anwendung konfirmatorischer Analysen zur Identifikation von Meßmodellen für einzelne Konstrukte zu kommen, führte zu den in Tabelle 27 dargestellten kongenerischen Subskalen. Die jeweils angekreuzten Items in einer Zeile konstituieren dabei eine gemeinsame Subskala. Die Tabelle enthält in der letzten Spalte auch die mittleren Reliabilitäten. Die dort angegebene Numerierung der Subskalen entspricht nicht notwendig der Numerierung der Faktoren in Tabelle 19, sondern wurden entsprechend dem Vorhandensein oder Nicht-Vorhandensein der extrinsischen, intrinsischen und sozialen Kernitems (vgl. S.33) zusammengestellt.

Tab. 27: Kongenerische Subskalen für die Magnitude- und kategorischen Messungen in beiden Nationen

SUBSKALA	1	2	3	4	5	6	7	8	9	10	11	12	13	RELIA
	SICH	EINK	AUFS	ACHT	FREI	INTE	SELB	VERA	KONT	HELF	MUTZ	SINN	BEDI	
ZUMA-KAT1	XXXX	XXXX	XXXX	XXXX										.534
ZUMA-MAG1	XXXX	XXXX	XXXX	XXXX										.704
NORC-KAT1	XXXX	XXXX	XXXX	XXXX										.456
NORC-MAG1	XXXX	XXXX	XXXX	XXXX										.640
ZUMA-KAT2						XXXX	XXXX	XXXX	XXXX					.480
ZUMA-MAG2					XXXX	XXXX	XXXX		XXXX			XXXX	XXXX	.647
NORC-KAT2						XXXX	XXXX	XXXX				XXXX		.525
NORC-MAG2		XXXX				XXXX	XXXX	XXXX				XXXX		.802
ZUMA-KAT3										XXXX	XXXX	XXXX	XXXX	.333
ZUMA-MAG3								XXXX		XXXX	XXXX			.656
NORC-KAT3									XXXX	XXXX	XXXX	XXXX		.592
NORC-MAG3									XXXX	XXXX	XXXX		XXXX	.790

Die Anpassungsindizes, Einzelladungen, Meßfehler der einzelnen Items und Einzelreliabilitäten sind Tabelle 28 (für die ZUMA-Stichprobe) und Tabelle 29 (für die NORC-Stichprobe) zu entnehmen.

Tab. 28: Einzelergebnisse der kongerischen Subskalen (ZUMA)

		B_i	θ_i^2	ρ_i	Anpassungsindizes
<u>ZUMA/KAT1</u>	SICHER	1.053	1.600	.409	$\chi^2 = .860$ (df=2; P= .656)
	EINKOM	1.259	1.332	.543	$\chi^2/df = .430$
	AUFSTG	1.276	.468	.776	GFI= .995
	GEACHT	1.093	1.750	.406	RMR= .041
<u>ZUMA/KAT2</u>	INTER	.476	.771	.227	$\chi^2 = 1.880$ (df=2; P= .390)
	SELBST	.941	.480	.648	$\chi^2/df = .940$
	VERANT	1.105	.535	.695	GFI= .991
	KONTKT	.853	2.017	.348	RMR= .036
<u>ZUMA/KAT3</u>	HELFEN	1.505	.924	.314	$\chi^2 = 4.850$ (df=2; P= .088)
	NUETZL	1.279	1.468	.527	$\chi^2/df = 2.425$
	SINNVL	.876	1.210	.388	GFI= .986
	BEDING	.482	2.036	.102	RMR= .106
<u>ZUMA/MAG1</u>	SICHER	.332	.056	.663	$\chi^2 = 1.260$ (df=2; P= .532)
	EINKOM	.291	.056	.603	$\chi^2/df = .630$
	AUFSTG	.365	.019	.886	GFI= .994
	GEACHT	.349	.062	.667	RMR= .002
<u>ZUMA/MAG2</u>	FRZEIT	.267	.070	.507	$\chi^2 = 11.200$ (df=9; P= .076)
	INTER	.257	.009	.880	$\chi^2/df = 1.240$
	SELBST	.223	.023	.685	GFI= .964
	KONTKT	.245	.063	.488	RMR= .003
	SINNVL	.266	.021	.741	
	BEDING	.260	.048	.578	
<u>ZUMA/MAG3</u>	VERANT	.258	.045	.598	$\chi^2 = .000$ (df=0; P=1.000)
	HELFEN	.310	.045	.681	
	NUETZL	.340	.052	.690	

Es sind in erster Linie drei Befunde, die bei der kongenerischen Subskaleneubildung Beachtung verdienen:

- Die Reliabilitäten der Magnitude-Subskalen sind in der Regel beträchtlich größer als die der kategorischen Skalen. Dieses Ergebnis, daß sich bei einem Vergleich der beiden Meßmethoden regelmäßig einstellt - auch bei der Untersuchung anderer inhaltlicher Stimuli als der Berufswerte - bringt zum Ausdruck, daß Magnitude-Skalen einen höheren Anteil wahrer Varianz der Gesamtvarianz abbilden (vgl. auch WEGENER 1984). Die Erklärung für diesen Umstand dürfte

Tab. 29: Einzelergebnisse der kongenerischen Subskalen (NORC)

		β_i	θ_i^2	ρ_i	Anpassungsindizes
<u>NORC/KAT1</u>	SICHER	.873	1.303	.369	$\chi^2 = 5.150$ (df=2; P= .076)
	EINKOM	1.058	.909	.552	$\chi^2/df = 2.575$
	AUFSTG	.977	1.047	.477	GFI= .978
	GEACHT	1.022	1.402	.426	RMR= .082
<u>NORC/KAT2</u>	INTER	.720	.609	.460	$\chi^2 = 2.840$ (df=2; P= .242)
	SELBST	.810	1.161	.361	$\chi^2/df = 1.420$
	VERANT	1.056	.221	.835	GFI= .986
	SINNVL	.708	.626	.445	RMR= .037
<u>NORC/KAT3</u>	SICHER	.993	1.194	.459	$\chi^2 = 6.940$ (df=5; P= .225)
	KONTKT	1.105	1.011	.547	$\chi^2/df = 1.388$
	HELFEN	.942	.193	.821	GFI= .975
	NUETZL	1.054	.443	.715	RMR= .045
	SINNVL	.690	.652	.422	
<u>NORC/MAG1</u>	SICHER	.263	.123	.360	$\chi^2 = .470$ (df=2; P= .789)
	EINKOM	.308	.090	.514	$\chi^2/df = .235$
	AUFSTG	.399	.021	.883	GFI= .989
	GEACHT	.424	.044	.804	RMR= .002
<u>NORC/MAG2</u>	EINKOM	.324	.080	.568	$\chi^2 = 11.910$ (df=5; P= .036)
	INTER	.404	.013	.926	$\chi^2/df = 2.382$
	SELBST	.365	.043	.756	GFI= .963
	VERANT	.395	.018	.897	RMR= .002
	SINNVL	.381	.023	.863	
<u>NORC/MAG3</u>	SICHER	.329	.084	.563	$\chi^2 = 12.860$ (df=5; P= .025)
	KONTKT	.422	.024	.881	$\chi^2/df = 2.570$
	HELFEN	.376	.014	.909	GFI= .954
	BEDING	.329	.030	.783	RMR= .004
	NUETZL	.381	.033	.815	

in der möglichen Differenzierbarkeit der Magnitude-Reaktionen gegenüber den kategorischen limitierten Antwortmöglichkeiten zu suchen sein. Insbesondere Extremurteile führen bei Kategorialmessungen zur Konfundierung der Endkategorien. Hier mag ein möglicher Grund dafür liegen, daß kategorisch gemessene Indikatoren ihre latenten Konstrukte häufig nicht 'rein' messen.

- Als zweites bemerkenswertes Ergebnis ist auf die Invarianz der Subskala hinzuweisen, die aus den vier extrinsischen Items 'Sichere Berufsstellung',

'Hohes Einkommen', 'Gute Aufstiegsmöglichkeiten' und 'Ein Beruf, der anerkannt und geachtet wird' besteht. Diese Subskala stellt gewissermaßen den Kern aller Messungen dar. Sie kommt ihrem Inhalt nach der generalisierten Oben/Unten-Dimension der PINDIS-Konfiguration am nächsten. Die Invarianz dieser Skala gegenüber einer Veränderung der Nationen und Methoden darf jedoch nicht darüber hinwegtäuschen, daß es sowohl bezüglich der Nationen als auch bezüglich der Methoden gewichtige Verschiebungen in der Verteilung der Reliabilitätswerte gibt; d.h. die Items messen die Konstrukte bei unterschiedlichen Methoden und Nationen unterschiedlich gut.

- Bei den Subskalen KAT2, KAT3, MAG2 und MAG3 ergibt sich ein mehr uneinheitliches Bild. Zwar zeichnen sich KAT2 und MAG2 bzw. KAT3 und MAG3 über die Nationen hinweg durch invariante Kernitems aus - 'Tätigkeit, bei der man selbständig arbeiten kann' und 'Interessante Tätigkeit' im Falle von KAT2 und MAG2 sowie 'Ein Beruf, bei dem man anderen helfen kann' und 'Ein Beruf, der für die Gesellschaft nützlich ist' im Falle von KAT3 und MAG3 - doch werden die Konstrukte über diese Kernitems hinaus bei den verschiedenen Nationen und Methoden in unterschiedlicher Weise weiter ausdifferenziert. Insgesamt steht fest, daß die Berücksichtigung der kongenerischen Meßmodelle sowohl für den kategorischen Fall als auch für den Magnitude-Fall in beiden Populationen die einfache und eingängig zu interpretierende Lösung (vgl. auch SCHMIDT 1984) nicht replizieren kann. Vielmehr finden sich unter diesem Gesichtspunkt methoden- und nationenspezifische Dimensionen. Angeleitet von den Einsichten der PINDIS-Rekonstruktion der Gruppenunterschiede kann allerdings festgestellt werden, daß der Übergang von den kategorischen zu den Magnitude-Messungen zu einer Generalisierung der Iteminhalte führt, und zwar in beiden nationalen Stichproben; d.h. die beiden zusätzlichen Skalen werden unspezifischer, wenn auch jeweils inhaltlich verschieden: Bei der ZUMA-Stichprobe erfolgt die Generalisierung in Hinblick auf 'intrinsische' bzw. 'sozial bezogenen' Items, während bei der NORC-Stichprobe deutlich auch 'extrinsische' Wertvorstellungen miteinbezogen werden. Für die Interpretation der Magnitude-Messungen bedeutet das, daß ein allgemeiner, über beide Populationen vergleichbarer Faktor nur in bezug auf die vier Items der Subskalen KAT1 und MAG1 einen identischen Inhalt hat, während die übrigen Items unterschiedliche Konstrukte repräsentieren; d.h. eine Tendenz zur 'Eindimensionalisierung', die mit der Magnitude-Methode gegeben ist, darf nicht zu dem Schluß verleiten, daß die gemeinsame Dimension auch

bedeutungsgleich ist. Erst die überaus strengen Anpassungskriterien der kongenerischen Einzelanalysen konnten dieses Ergebnis, das sich aus den Ergebnissen der exploratorischen Analysen nicht so ohne weiteres herauslesen ließ, hervorbringen.

Für die kategorisch gemessenen Items der beiden Populationen ist die Spannweite der reliablen Items entsprechend kleiner, aber interkulturell ebenfalls nicht bedeutungsgleich. Während das Item 'Gibt mir das Gefühl, etwas Sinnvolles zu tun' in der NORC-Stichprobe bei beiden Subskalen auftaucht und 'Viel Kontakte zu anderen Menschen' zur Skala KAT3 zählt, beinhaltet KAT3 bei der ZUMA-Stichprobe das Item 'Gesunde Arbeitsbedingungen', und 'Viel Kontakte zu anderen Menschen' gehört zu KAT2.

5.5 Zur Vorhersage von Gruppenunterschieden

Im Zusammenhang mit der Vorhersage von Gruppenunterschieden interessierten uns in erster Linie die folgenden beiden Fragestellungen:

- Welche Berufswerte und Dimensionen sind für die Unterscheidung der ZUMA-Stichprobe und der NORC-Stichprobe besonders indikativ;
- Welche der in Tabelle 19 (S.59) dargestellten Faktoren differenzieren zwischen den Stufen wichtiger vergleichbarer sozialer Gruppierungsmerkmale in ähnlicher Weise.

Die erste Frage wurde mit Hilfe diskriminanzanalytischer Techniken zu beantworten versucht. Für die kategorischen und Magnitude-Messungen wurden getrennte stufenweise Diskriminanzanalysen mit den beiden Nationen als Gruppen gerechnet. Es ergaben sich die in Tabelle 30 dargestellten (standardisierten) Koeffizienten der beiden kanonischen Diskriminanzfunktionen. Die Variablen, für die in der Tabelle kein Wert eingetragen ist, fielen bei der stufenweisen Analyse heraus.

Betrachtet man zunächst die Ergebnisse für den Magnitude-Fall, so fällt auf, daß die Diskriminanzfunktion eine Dimension 'intrinsisch'-'nicht intrinsisch' darzustellen scheint: Alle genuin intrinsischen Items korrelieren mit dieser Funktion negativ. Im kategorischen Fall ist dies nicht so ausgeprägt, doch fallen

auch hier die negativen oder niedrigen Ladungen der intrinsischen Items auf.

Tab. 30: Koeffizienten der kanonischen Diskriminanzfunktionen
(standardisiert) zur Unterscheidung ZUMA/NORC

	<u>KATEGORISCH</u>	<u>LOG MAGNITUDE</u>
SICHER	-----	-----
EINKOM	-----	.321
AUFSTG	.477	.679
GEACHT	-----	-----
FRZEIT	-.176	-.403
INTER	.268	-.288
SELBST	-.275	-.340
VERANT	.341	-.288
KONTKT	-----	-.285
HELFEN	.241	.778
NUETZL	.256	-----
SINNVL	-----	-----
BEDING	.142	-----

Auffällig ist auch die hohe Ladung des Items 'Ein Beruf, bei dem man anderen helfen kann' auf der Magnitude-Diskriminanzfunktion.

Zur Beantwortung der zweiten oben genannten Frage wurden für die einzelnen Befragten Faktorenwerte berechnet. Die Mittelwerte der Faktorenwerte für die Stufen einiger wichtiger sozialer Variablen wurden dann dahingehend untersucht, ob sie über die Nationen und die Methoden hinweg vergleichbaren Trends gehorchen.

Im Falle von ZUMA wurden die Variablen Alter, Geschlecht, Bildung, Partnererfahrung (schon einmal mit einem Partner zusammengelebt oder nicht), Erwerbstätigkeit (erwerbstätig vs. nebenher- oder nicht erwerbstätig), Schichtzugehörigkeit, Selbsteinstufung auf der Oben-Unten-Skala, Selbsteinstufung auf der Rechts-Links-skala betrachtet, im Falle von NORC die Variablen Alter, Geschlecht, Bildung,

Partnererfahrung, Schicht und Liberal/Konservativ.

Leider ergab sich nur im Fall der ZUMA-Stichprobe und hier auch nur für den extrinsischen Faktor (F1 bei Kategorisch und F3 bei Magnitude) ein einheitliches Bild. Tabelle 31 zeigt, daß in diesem Fall die kategorischen und die Magnitude-

Tab. 31: Vergleich Kategorialskalierung/Magnitude-Messung (ZUMA)

	<u>ALTER</u>		
	<u>18-20</u>	<u>31-45</u>	<u>46-65</u>
KATEGORISCH	-.374	.213	.414
LOG MAGNITUDE	-.261	.067	.399

	<u>GESCHLECHT</u>	
	<u>Männlich</u>	<u>Weiblich</u>
KATEGORISCH	-.080	.069
LOG MAGNITUDE	.010	-.030

	<u>BILDUNG</u>		
	<u>Niedrig</u>	<u>Mittel</u>	<u>Hoch</u>
KATEGORISCH	.427	-.050	-.407
LOG MAGNITUDE	.324	-.123	-.158

	<u>PARTNERERFAHRUNG</u>	
	<u>Ja</u>	<u>Nein</u>
KATEGORISCH	.121	-.293
LOG MAGNITUDE	.139	-.329

	<u>ERWERBSTÄTIGKEIT</u>	
	<u>Erwerbstätig</u>	<u>Nebenher bzw. Nicht erwerbstätig</u>
KATEGORISCH	-.056	.288
LOG MAGNITUDE	-.085	.569

	<u>RECHTS/LINKS</u>	
	<u>Rechts</u>	<u>Links</u>
KATEGORISCH	.472	-.521
LOG MAGNITUDE	.519	-.417

Messungen für die dort abgebildeten Variablen einem vergleichbaren monotonen Trend folgen. Die Ergebnisse erscheinen darüber hinaus auch inhaltlich sehr plausibel. Leider ließen sich für den Methodenvergleich innerhalb von NORC und in bezug auf einen Vergleich über die Nationen hinweg keine derartig plausiblen Ergebnisse finden.

6. ZUSAMMENFASSUNG UND SCHLUSSFOLGERUNGEN

Die Untersuchung und ihre Befunde thematisieren das Problem der Vergleichbarkeit im Grunde aus unterschiedlichen Perspektiven. Eine Zusammenfassung wird pragmatische, meßtheoretische und substantielle Gesichtspunkte berücksichtigen müssen.

Die pragmatische Perspektive führt zu der Feststellung, daß ein Methodenvergleich unter feldmäßigen Bedingungen und in unterschiedlichen Ländern durchaus möglich ist - mit allen Einschränkungen, die die Notwendigkeit einer Kooperation auch in dieser Studie offenbart hat. Die vorliegende Untersuchung läßt Aussagen über Wirkung von Meßmethoden zu, obwohl die Stichproben nicht in der wünschenswerten Weise als Parallelstichproben realisiert werden konnten und ein vollständig einheitliches Forschungsprogramm letztendlich nicht realisiert werden konnte. Hervorzuheben ist, daß der Einsatz der feldmäßig schwierigen Magnitude-Methode - zumal an inhaltlich komplexen Stimuli - in beiden Fällen gelang. Allerdings ist die Realisierung in den USA, wo die Methode in dem Format, das bei ZUMA ausgetestet wurde, erstmalig eingesetzt wurde, verbesserungsbedürftig.

Unter meßtheoretischen Gesichtspunkten sind zwei Resultate besonders hervorzuheben: der Nachweis der wesentlichen Einflüsse der verschiedenen Meßverfahren auf das Antwortverhalten und der Nachweis unterschiedlich hoher Reliabilitäten bei den hier verwendeten Verfahren der Kategoriale- und Magnitude-Skalierung.

Für kategoriale Skalierungen zeigte sich, daß numerische und verbale Versionen nicht naiv und unreflektiert durcheinander ersetzt werden können. Hier ist vielmehr mit Einflüssen des unterschiedlichen semantischen Verständnisses der verbalen Antwortvorgaben zu rechnen. Die Untersuchung der für die Antwortvorgaben relevanten adverbialen Modifikatoren zeigte, daß es bedeutsame Unter-

schiede im Verständnis der Antwortvorgaben in bezug auf sozio-demographisch klassifizierte Gruppen gibt. Mit Einflüssen des unterschiedlichen semantischen Verständnisses ist aber nicht nur reaktionsseitig zu rechnen. Vielmehr zeigen die Ergebnisse, daß ein unterschiedliches Verständnis quantitativer Stimulusbegriffe ('sicher' in 'Sichere Berufsstellung') in der Lage ist, durch Wechselwirkung mit einem unterschiedlichen Verständnis der Antwortvorgaben sozio-demographische Gruppenunterschiede zu nivellieren.

Der zweite meßtheoretische Befund, die eindeutig höheren Reliabilitäten der Magnitude-Messungen, dürfte im wesentlichen mit dem größeren Range und den größeren Differenzierungsmöglichkeiten der Magnitude-Reaktionen zusammenhängen.

Im übrigen konnte eine Überlegenheit der Magnitude-Skalen im Sinne erfüllter Meßstrukturen aufgrund der offensichtlich vorliegenden Mehrdimensionalität der Items nur für die vier extrinsischen Items nachgewiesen werden. Bei den axiomatischen Überprüfungen zeigte sich aber auch, daß die numerische Kategorienskala in bezug auf diese vier Items die Qualität einer Intervallskala hatte. Weitere theoretisch begründete Hinweise auf die Skalenqualität ergab die Überprüfung der Urteilsmodelle. Sie zeigte, daß die kategorialen und Magnitudemessungen in Übereinstimmung mit den entsprechenden urteilstheoretischen Grundlagen stehen. Leider konnte diese Analyse nur auf aggregiertem Niveau und - ebenso wie die axiomatischen Überprüfungen - nur im Kontext der deutschen Teilstudie durchgeführt werden. In beiden Nationen konnte lediglich die numerische Skalenqualität anhand der Intermodalitäten-Beziehungen und der Interskalen-Relationen überprüft werden. Dabei zeigte sich, daß die amerikanischen Magnitude-Messungen von minderer Qualität waren als die deutschen, wofür sich eine Reihe einleuchtender Erklärungen (vgl. S.51) anboten.

Die substantielle Perspektive betrifft die eigentliche inhaltliche Frage, die mit dieser Untersuchung gestellt wurde. Letztere bezog sich auf die Entscheidung der Vergleichbarkeit der beruflichen Einstellungen von zwei Stichproben unterschiedlicher Nationalität und damit auf das eigentliche Validitätsproblem. Nur zu deutlich dürfte sein, daß diesbezüglich nur eine exemplarische, paradigmatische Antwort möglich ist. Auch unabhängig von den Einschränkungen der konkreten Realisierung lassen die kleinen Befragtengruppen keine allgemeinen Aussagen zu. Unter diesem Vorbehalt stellt der aufschlußreichste Befund dieser Untersuchung

allerdings die Heterogenität der gefundenen Strukturen dar. Sie belegt eindrucksvoll, daß der numerische Vergleich von bloßen Skalenwerten über Kultur- und Ländergrenzen hinweg ohne Aussagekraft ist. Wir finden für die 13 Berufswerte in den amerikanischen und deutschen Populationen keine einheitliche Konstruktstruktur, die den Übertragungsschluß auch inhaltlicher Identität zuließe. Allerdings hat die Auswertung der Daten auf zwei Wegen zu der Möglichkeit geführt, die Art der Verschiedenheit näher zu bestimmen. Der Einsatz der zwei unterschiedlichen Meßverfahren bei den Erhebungen bot dafür die Voraussetzung.

Unter Einschluß beider Erhebungsmethoden und beider Nationen konnte mit Hilfe der multidimensionalen PINDIS-Analyse ein gemeinsamer 'Bewertungsraum' rekonstruiert werden. Während nun für die kategorischen Messungen dieser Bewertungsraum in seinen Dimensionen lediglich mit unterschiedlichen Gewichten versehen werden muß, gilt für den Einsatz der Magnitude-Methodologie, daß es zu einer Vereinheitlichung kommt. In beiden Untersuchungsstichproben führt die Methode zu einer Reduktion der Dimensionalität und damit offenbar zu einer Vergleichsdimension für die beiden Populationen. Freilich geht dieser Prozeß zu Lasten jener Gesichtspunkte, die im kategorischen Fall auf getrennten Dimensionen abgebildet werden, und hier verhalten sich die beiden nationalen Stichproben verschieden hinsichtlich der Items, die gewissermaßen in die allgemeine Dimension integriert werden. Die Reduktion impliziert daher keineswegs schon - jedenfalls für die hier untersuchten komplexen mehrdimensionalen Stimuli - Vergleichbarkeit.

Die beschriebene Konstruktreduktion beim Übergang zur Magnitude-Skalierung deutet sich schon über die Anteile der Varianzaufklärungen der einzelnen Faktoren an, die sich bei den exploratorischen Faktorenanalysen ergeben haben. Läßt man neben dem die weitaus meiste Varianz aufklärenden Hauptfaktor noch die übrigen Faktoren zu, so läßt sich bereits vermuten, daß man bei Anlegen entsprechend strenger Kriterien auch Indikatoren bzw. kongenerische Meßmodelle in bezug auf diese Faktoren finden kann. Bei der Suche nach möglichen Items, die in bezug auf die entsprechenden Konstrukte der exploratorischen Faktorenanalysen kongenerische Meßmodelle bilden, zeigte sich, daß streng kongenerische Subskalen, die über die Methoden und Nationen hinweg aus identischen Items bestehen, nur für vier der 13 Items nachweisbar sind. Diese vier Items bilden einen invarianten vergleichbaren Kern der Itembatterie. Es handelt sich um die vier extrinsischen Items 'Sichere Berufsstellung', 'Hohes Einkommen', 'Gute Aufstiegsmöglichkeiten'

und 'Ein Beruf, der anerkannt und geachtet wird'. Diese Items sind deutlich statusgeprägte und hierarchiebezogene Berufsmerkmale. Man kann unterstellen, daß diese Items unabhängig von der Meßmethode Vergleichbarkeit sichern und daß sie für komparative Untersuchungen brauchbar sind.

Die Aufnahme anderer Items als der genannten vier ist dagegen problematischer. Mit ihnen werden uneinheitliche, methoden- und populationsabhängige Konstrukte gemessen. Während diese Konstrukte im kategorischen Fall jeweils spezifisch sind (d.h. nur wenige Items umfassen), sind sie für die Magnitude-Messungen genereller. Diese entspricht den unterschiedlichen Wirkungen der Methoden. Kategorialskalierungen führen zu klarer geschnittenen Faktorenstrukturen und einer größeren Anzahl von Faktoren als Magnitude-Messungen. Aber in beiden Fällen sind in bezug auf die resultierenden Konstrukte kaum vergleichbare Aussagen möglich, sieht man von dem Vorhandensein von jeweils zwei Kernitems ab.

Bei der Bewertung der Subskalenbildung ist allerdings darauf hinzuweisen, daß die Etablierung der Skalen als 'reine' Skalen extrem strenge Anforderungen gestellt hat, Anforderungen, die bei weitem über die Standards, die in der Umfrageforschung üblich sind, hinausgehen. Bei Lockerung dieser Kriterien (Abfinden mit schlechteren Anpassungen der Meßmodelle) würden die Skalen allgemeiner, es ließen sich zusätzliche Items integrieren, und es käme zu stärkeren Überlappungen der Skalen. In diesem Sinne ist zu erwarten, daß eine Magnitude-Messung der Berufswerte einen größeren Anteil der Items kongenerisch messen kann - die Reliabilitäten dieser Skalierungsart liegen erheblich höher als die von kategorischen Messungen. Im Sinne der multidimensionalen Konfigurationsskalierung könnte man also für einen internationalen Vergleich deutlich mehr Items benutzen, wenn man diese Items magnitudeskaliert. Allerdings sollte sich nach unseren Analysen ein solcher Versuch nicht auf alle 13 Berufswerte beziehen. Welche der Items - über die vier extrinsischen Items hinaus - für diesen Zweck brauchbar ist, kann nur eine Untersuchung an einem größeren Datenmaterial zeigen.

Haben die direkten Nationenvergleiche - zumindest was die nicht-extrinsischen Items angeht - keine hinreichend scharf umrissenen Ergebnisse gebracht, so ergeben wenigstens entsprechende Diskriminanzanalysen einige konkrete Hinweise darüber, bezüglich welcher Item-Dimension sich die Nationen unterscheiden. Die unterscheidende Dimension scheint wesentlich durch die beiden Pole 'intrinsisch'

'nicht-intrinsisch' charakterisiert zu sein.

Ein Vergleich zwischen den Nationen als Ganzen vernachlässigt noch, daß es innerhalb der Nationen vergleichbare soziodemographisch differenzierbare Untergruppen geben kann. Sollen die gefundenen Dimensionen bzw. Konstrukte in irgendeiner Weise über die Nationen und Methoden hinweg soziale Gruppenuntergruppen in einheitlicher und invarianter Weise kennzeichnen, so wären entsprechende einheitliche Trends bezüglich der für die einzelnen Gruppen bestimmten mittleren Faktorenwerte zu erwarten gewesen. Leider konnten derartige Trendähnlichkeiten nur für die extrinsischen Kategorial- und Magnitundefaktoren und nur im Fall der ZUMA-Stichprobe nachgewiesen werden. Immerhin ist auch dies - wenn auch mit Einschränkungen - ein weiterer Hinweis auf die Stabilität der extrinsischen Skala.

LITERATUR

- AHMAVAARA, Y. The mathematical theory of factorial invariance under selection. Psychometrika, 1954a, 19, 27-38
- AHMAVAARA, Y. Transformational analysis of factorial data. Annales Academiae Scientiarum Fennicae, 1954b, 88, 1-150
- ALLERBECK, K. Analysis and inference in cross-cultural research. In: A. Szalai & R. Petrella (Eds.), Cross-National Comparative Survey Research. New York: Pergamon Press, 1977
- ALWIN, D.F. & JACKSON, D.J. Measurement models for response errors in surveys: Issues and applications. In: K. Schuessler (Ed.), Sociological Methodology 1980. San Francisco: Jossey Bass, 1980
- BARNES, S., KAASE, M. et al. Political Action. Beverly Hills: Sage, 1979
- BECK, W., BRATER, M. & WEGENER, B. Berufswahl und Berufszuweisung. Zur sozialen Verwandtschaft von Ausbildungsberufen. Frankfurt: Campus, 1979
- BOOMSMA, A. The robustness of LISREL against small sample sizes in factor analysis models. In K.G. JÖRESKOG & H. WOLD (Eds.), Systems under Indirect Observations. Part I. Amsterdam: North-Holland, 1982
- BORG, I. Geometric representation of individual differences. In: J.C. Lingoes (Ed.), Geometric Representations of Relational Data. Readings in Multidimensional Scaling. Ann Arbor, Michigan: Mathesis Press, 1977
- BORG, I. & LINGOES, J.C. What weight should we have in individual difference scaling? Quality and Quantity, 1978, 12, 223-237
- BRISLIN, R.W. Cross-cultural research in psychology. Annual Review of Psychology, 1983, 34, 363-400
- BRISLIN, R.W., LONNER, W. & THORNDIKE, R. Cross-Cultural Research Methods. New York: John Wiley, 1973
- CARMINES, E.G. & McIVER, J.P. Analyzing models with unobserved variables. In: G.W. Bohrnstedt & E.F. Borgatta (Eds.), Social Measurement: Current Issues. Beverly Hills: Sage, 1981
- CARROLL, J.D. & WISH, M. Models and methods for three-way multidimensional scaling. In: D.H. Krantz, R.C. Atkinson, D.R. Luce & P. Suppes (Eds.), Contemporary Developments in Mathematical Psychology. Vol II. San Francisco: Freeman, 1974
- CATTELL, R.B. Personality and Motivation Structure and Measurement. Yonkers-on-Hudson: World Book, 1957

- CLIFF, N. Adverbs as multipliers. Psychological Review, 1959, 66, 27-44
- CLIFF, N. Adverb-adjective combination in overseas groups. Research Memorandum Nr. 16, Princeton, N.J.: Educational Testing Service, 1960
- ELDER, J.W. Comparative cross-national methodology. Annual Review of Sociology, 1976, 2, 209-230
- FAULBAUM, F. Konfirmatorische Analysen der Reliabilität von Wichtigkeitseinstufungen beruflicher Merkmale, Zumanachrichten, 1983, 13, 22-44
- FAULBAUM, F. Zur Konstruktioninvarianz numerischer und verbaler Kategorienskalen. Zumanachrichten, 1984, 14, 46-59
- GRUNERT, K.G. Die Eignung des Magnitude-Verfahrens zur Skalierung der Intensität des Empfindens von Kaufrisiken. Arbeitspapier Nr. 9, Institut für Haushalts- und Konsumökonomik, Universität Hohenheim, Stuttgart-Hohenheim, 1980
- GRUNERT, K.G. Die Ermittlung entscheidungsrelevanter Produktmerkmale beim Automobilkauf. In: E. Dichtl, H. Raffee & V. Potucek (Hg.), Marktforschung im Automobilsektor. Frankfurt: Verband der Automobilindustrie, 1983a
- GRUNERT, K.G. Magnitude-Skalierung. Marketing-ZFP, 1983b, 5, 108-112
- HARMAN, H.H. Modern Factor Analysis. Chicago: Chicago University Press, 1967²
- HOWE, E.S. Probabilistic adverbial qualifications of adjectives. Journal of Verbal Learning and Verbal Behavior, 1962, 1, 225-242
- JACOB, B.M. & JACOB, P.E. Leader's values and community activeness: The international studies of values in politics. In: A. Szalai & R. Petrella (Eds.), Cross-National Comparative Survey Research. New York: Pergamon Press, 1977
- JÖRESKOG, K.G. Structural analysis of sets of congeneric tests. Psychometrika, 1971, 36, 109-133
- JÖRESKOG, K.G. Structural analysis of covariance and correlation matrices. Psychometrika, 1978, 43, 443-477
- JÖRESKOG, K.G. Statistical estimation of structural models in longitudinal-developmental investigations. In: J.R. Nesselroade & P.B. Baltes (Eds.), Longitudinal Research in the Study of Behavior and Development. New York-London: Academic Press, 1979
- JÖRESKOG, K.G. & SÖRBOM, D. Statistical models and methods for test-retest situations. In: D.N.M. de Gruijter & L.J.T. Van der Kamp (Eds.), Advances in Psychological and Educational Measurement. New York-London: John Wiley, 1976
- JÖRESKOG, K.G. & SÖRBOM, D. EFAP. Exploratory Factor Analysis Program. User's Guide. Version II. National Educational Resources, Inc., 1978.

- JÖRESKOG, K.G. & SÖRBOM, D. LISREL V. User's Guide. National Educational Resources, Inc, 1981
- KENDALL, M. & STUART, A. The Advanced Theory of Statistics. Vol. II. London-High Wycombe: Charles Griffin, 1979⁴
- KRISTOF, W. Das Cliffsche Gesetz im Deutschen. Psychologische Forschung, 1966, 29, 22-31
- KRUSKAL, J.B. Nonmetric multidimensional scaling: A numerical method. Psychometrika 1964, 29, 1-27, 115-129
- LILLY, R.S. The qualification of evaluative adjectives by frequency adverbs. Journal of Verbal Learning and Verbal Behavior, 1968a, 7, 333-336
- LILLY, R.S. Multiplying values of intensive probabilistic, and frequency adverbs when combined with potency adjective. Journal of Verbal Learning and Verbal Behavior, 1968b, 7, 854-858
- LINGOES, J.C. & BORG, I. Optimal solutions for dimension and vector weights in PINDIS. Michigan Mathematical Psychology Program, 1976, 4, 1-18
- NAROLL, R. Some thoughts on comparative method in cultural anthropology. In: H.M. Blalock & A.B. Blalock (Eds.), Methodology in Social Research. New York: McGraw-Hill, 1968
- NASH-PARKER, R. Measuring social participation. American Sociological Review, 1983, 48, 864-873
- NOWAK, S. The strategy of cross-national survey research for the development of social theory. In: A. Szalai & R. Petrella (Eds.), Cross-National Comparative Survey Research. New York: Pergamon Press, 1977
- ORTH, B. Einführung in die Theorie des Messens. Stuttgart: Kohlhammer, 1974
- ORTH, B. A theoretical and empirical study of scale properties of magnitude estimation and category rating scales. In: B. Wegener (Ed.), Social Attitudes and Psychophysical Measurement. Hillsdale, N.J.: Erlbaum, 1982
- ORTH, B. & WEGENER, B. Scaling occupational prestige by magnitude estimation and category-rating methods: A comparison with the sensory domain. European Journal of Social Psychology, 1983, 13, 417-431
- PRZEWORSKI, A. & TEUNE, H. Equivalence in cross-national research. Public Opinion Quarterly, 1966, 30, 33-43
- PRZEWORSKI, A. & TEUNE, H. The Logic of Comparative Social Inquiry. New York: Wiley-Interscience, 1970
- ROHRMANN, B. Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. Zeitschrift für Sozialpsychologie, 1978, 9, 222-2.

- RUMMEL, R.J. A field theory of social action with application to conflict within nations. General Systems, 1965, 10, 183-211
- RUMMEL, R.J. Applied Factor Analysis. Evanston: Northwestern University Press, 1970
- SARIS, W., NEIJENS, P. & VAN DOORN, L. Scaling social variables by multimodality matching. Methoden en Data Nieuwsbrief, 1980, 5, 3-21
- SCHEUCH, E.K. Society as a context in cross-national comparisons. Social Science Information, 1967, 6, 7-23
- SCHEUCH, E.K. The cross-cultural use of sample surveys: Problems of comparability. In: R. Rokkan (Ed.), Comparative Research across Cultures and Nations. Paris: Mouton, 1968
- SCHMIDT, P. Messung von Arbeitsorientierungen: Theoretische Fundierung und Test alternativer Meßmethoden. Analyse und Kritik, 1983, 2
- TEUNE, H. Measurements in comparative research. Comparative Political Studies, 1968, 1
- TEUNE, H. Analysis and interpretation in cross-national survey research. In: A. Szalai & R. Petrella (Eds.), Cross-National Comparative Survey Research. New York: Pergamon Press, 1977
- VERBA, S. The cross-national program in political and social change: A history and some comments. In: A. Szalai & R. Petrella (Eds.), Cross-National Comparative Survey Research. New York: Pergamon Press, 1977
- VERBA, S., NIE, N.H. & KIM, J.-O. Participation and Political Equality. Cambridge: Cambridge University Press, 1978
- WEGENER, B. Einstellungsmessung in Umfragen: Kategorische vs. Magnitude-Skalen. Zumanachrichten, 1978, 3, 3-27
- WEGENER, B. Magnitude-Messung in Umfragen: Kontexteffekte und Methode. Zumanachrichten, 1980, 6, 4-40
- WEGENER, B. Category and Magnitude Scales of Multidimensional Importance Judgments: A NORC-ZUMA Comparison. Intermediate Report given at Havard University, Cambridge, USA, September 24, 1982a
- WEGENER, B. Fitting category to magnitude scales for a dozen survey-assessed attitudes. In: B. Wegener (Ed.), Social Attitudes and Psychophysical Measurement. Hillsdale, N.J.: Erlbaum, 1982b
- WEGENER, B. Zwischenbericht des Projekts 'Skalenvergleich'. ZUMA-Arbeitsbericht, Mannheim, 1983a

- WEGENER, B. Category-rating and magnitude estimation scaling techniques. Sociological Methods and Research, 1983b, 12, 31-75
- WEGENER, B. Ergebnisse der Reliabilitätsstudie 1982. ZUMA-Arbeitsbericht, Mannheim 1984
- WEGENER, B. & KIRSCHNER, H.-P. A note on estimating interscale relations in 'direct' psychophysical scaling. British Journal of Mathematical and Statistical Psychology, 1981, 34, 194-204
- WEGENER, B., FAULBAUM, F. & MAAG, G. Die Wirkung von Antwortvorgaben bei Kategoriaskalen. Zumanachrichten, 1982a, 10, 3-20
- WEGENER, B., FAULBAUM, F. & MAAG, G. Die Wirkung adverbialer Modifikatoren. Psychologische Beiträge, 1982b, 24, 343-345
- WHITING, J.W.M. Methods and problems in cross-cultural research. In: G. Lindzey & A. Aronson (Eds.), The Handbook of Social Psychology. Vol. II. Reading, Mass. Addison-Wesley, 1968

ZUMA-Arbeitsberichte

- 80/15 Gerhard Arminger, Willibald Nagl, Karl F. Schuessler
Methoden der Analyse zeitbezogener Daten. Vortragsskripten der ZUMA-Arbeitstagung vom 25.9.-5.10.79
- 81/07 Erika Brückner, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer, Peter Schmidt
Methodenbericht zum "Nationalen Sozialen Survey 1980"
- 81/19 Manfred Küchler, Thomas P. Wilson, Don H. Zimmerman
Integration von qualitativen und quantitativen Forschungsansätzen
- 82/03 Gerhard Arminger, Horst Busse, Manfred Küchler
Verallgemeinerte Lineare Modelle in der empirischen Sozialforschung
- 82/08 Glenn R. Carroll
Dynamic analysis of discrete dependent variables: A didactic essay
- 82/09 Manfred Küchler
Zur Messung der Stabilität von Wählerpotentialen
- 82/10 Manfred Küchler
Zur Konstanz der Recallfrage
- 82/12 Rolf Porst
"ALLBUS 1982" - Systematische Variablenübersicht und erste Ansätze zu einer Kritik des Fragenprogramms
- 82/13 Peter Ph. Mohler
SAR - Simple AND Retrieval mit dem Siemens-EDT-Textmanipulationsprogramm
- 82/14 Cornelia Krauth
Vergleichsstudien zum "ALLBUS 1980"
- 82/21 Werner Hagstotz, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer
Methodenbericht zum "ALLBUS 1982"
- 83/09 Bernd Wegener
Two approaches to the analysis of judgments of prestige: Interindividual differences and the general scale
- 83/11 Rolf Porst
Synopsis der ALLBUS-Variablen. Die Systematik des ALLBUS-Fragenprogramms und ihre inhaltliche Ausgestaltung im ALLBUS 1980 und ALLBUS 1982
- 84/01 Manfred Küchler, Peter Ph. Mohler
Qualshop (ZUMA-Arbeitstagung zum "Datenmanagement bei qualitativen Erhebungsverfahren") - Sammlung von Arbeitspapieren und -berichten, Teil I + II
- 84/02 Bernd Wegener
Gibt es Sozialprestige? Konstruktion und Validität der Magnitude-Prestige-Skala

- 84/03 Peter Prüfer, Margrit Rexroth
Beschreibung eines Verfahrens zur Bewertung von Interviewerverhalten
- 84/04 Frank Faulbaum
Ergebnisse der Methodenstudie zur internationalen Vergleichbarkeit von Einstellungsskalen in der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1982
- 84/05 Jürgen Hoffmeyer-Zlotnik
Wohnquartiersbeschreibung. Ein Instrument zur Bestimmung des sozialen Status von Zielhaushalten
- 84/06 Rolf Porst
ALLBUS-Bibliographie (3. Fassung, Stand: 30.06.84)